

Toward the development of a dynamic dashboard for FutureLearn MOOCs: insights and directions

Mahsa Chitsaz

m.chitsaz@unsw.edu.au

PVC (Education)

UNSW Australia

Lorenzo Vigentini

l.vigentini@unsw.edu.au

PVC (Education)

UNSW Australia

Andrew Clayphan

a.clayphan@unsw.edu.au

PVC (Education)

UNSW Australia

In recent years, many higher education institutions have invested in the development of Massive Open Online Courses (MOOCs). With the increase of available MOOC data, there is an opportunity to provide insights to educators and developers into learners' behaviors through learning analytics. Focusing on the FutureLearn platform (FL), standardized data files are offered to partner institutions. Additionally, a report is offered to stakeholders, but it is limited in a number of ways: it is static, it is limited in presenting relevant information and, most importantly, it does not provide 'real-time' access to data. This paper provides an overview of the rationale and the development process of a dynamic and near real-time dashboard. It explores the viability of different types of visualizations with the available data, lessons learned, comparisons with similar efforts, and future directions are discussed.

Keywords: FutureLearn MOOCs, Dashboards, Visualizations, Sense-making, Learning Analytics.

Introduction

Following the hype of MOOCs (Massive Open Online Courses) our institution entered the field in the consolidation phase of the hype cycle (Linden & Fenn, 2003) and invested in the experimentation with two very different platforms: Coursera and FutureLearn. Over the past two years UNSW developed and delivered over 20 MOOCs focusing on innovative pedagogy and variety over a mass-production approach. This was specifically to steer away from the media focus, which has been predominantly on a few characteristics of MOOCs – i.e. free courses, massive numbers, massive dropouts and implicit quality warranted by the status of the institutions delivering these courses. Instead, aligning with the rapidly growing research interest the idea was to question the effectiveness of MOOCs for learning and their pedagogies (Bayne & Ross, 2014; Yuan & Powell, 2013) and the possibility to take what we learnt in the process 'back to mainstream' of institutional practices. Questioning pedagogies, has driven the need to understand, analyze and evaluate MOOCs. Within an institution, this tends to rely on data analysts (Kandel, Paepcke, Hellerstein, & Heer, 2011) to interpret, translate and sometimes deduce academic and institutional requirements before modelling of data can take place, often before being able to deliver usable research outputs. Yet, despite the hype of big data in education and the potential associated with the ability to collect and analyze large amounts of information about students' learning behaviors, one of the biggest limitations has been finding a way to expose this data in a meaningful and relevant way to stakeholders. Visualization of data and the ability to manipulate visualizations in dashboards has been demonstrated to offer useful insights (Duval, 2011; Macfadyen & Dawson, 2010), but brought an over-expectation of benefits (Verbert et al., 2014). The focus of dashboards has been on providing awareness, reflection and sense-making, however, dashboards in and of themselves, do not automatically confer learning or awareness gains. It is a combination of effective design, requirements elicitation, and an understanding of user objectives, which can afford these products to aid data exploration and analysis processes.

Coursera provides a dashboard to educators and developers with a live view of their data, but the granularity of the information does not necessarily cater to the needs of all the interests from the different stakeholders. Others like edX, have analytical plug-in modules which users can install on their systems, affording similar functionalities (Cobos, Gil, Lareo, & Vargas, 2016; Fredericks, Lopez, Shnayder, Rayyan, & Seaton, 2016; Ruiz et al., 2014). There are also examples of visualization dashboards (Shi, Fu, Chen, & Qu, 2014) representing a range of behaviors from MOOCs, however a key question remains of whether information presented conveys useful insights. FutureLearn decided not to offer a fully-fledged dashboard to partner institutions. However, they explicitly chose to make simple datasets (in terms of granularity and stability, given that the datasets were designed independently from the inevitable changes and improvements to the user interface) available daily so that interested parties would be able to carry out their own analyses. Furthermore, they gave partner institutions a standard R script to generate pdf reports answering some commonly asked questions. These datasets provided both a challenge and an opportunity: in previous work a workflow was developed to manage analytics for

courses in the Coursera platform (Clayphan, Vigentini, Zhang, & Zhao, 2015). Working with FutureLearn data gave the opportunity to test the process with a different platform. Furthermore, taking advantage of the near real-time release of data, allowed for experimentation with different modes of representation and workflows that had not been possible with Coursera. Incidentally, the work presented here could be compared directly with a similar process developed independently, but around the same time, conducted by colleagues in Europe (Leon Urrutia, Cobos, Dickens, White, & Davis, 2016).

The paper is organized as follows: first the datasets are presented considering some key assumptions and limitations; then the methodology for building the dashboard is explained including some exemplars of the visualizations created; finally, issues and future directions are discussed.

Datasets

Each MOOC published in the FutureLearn website is presented in a hierarchical structure with weeks, activities and steps. Steps contain different types of material and can be recognized by the label next to the step title. In the courses designed at UNSW, eight different step types have been used: article, discussion, video, exercise, quiz, test, audio and LTI activity. Figure 1 shows an overview of a typical FutureLearn MOOC.

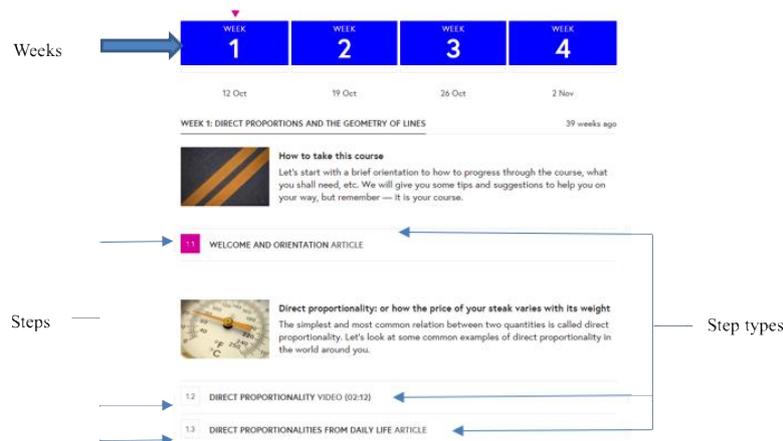


Figure 1: The Overview of a FutureLearn MOOC structure.

By default, FutureLearn provides eight data sources prepared as comma separated values (CSV) files. Table 1 describes each file with detail about their purpose. Conscious that MOOC platforms are relatively young and tend to rapidly develop features to improve the user experience, at the time of writing, the datasets available in FutureLearn have three main limitations: 1) granularity of user activity (currently limited to the time of the first/last access rather than a full interaction log); 2) minimal contextual information (lack of metadata about the learning context, such as video interaction data), and 3) partial demographic information to understand learners (only about 10% of participants have chosen to share personal details in the platform). As mentioned earlier, the limitations are the result of FutureLearn's choice to provide easy to access and stable datasets. These sources provide an excellent starting point to demonstrate the use of analytics in action.

Table 1: FutureLearn Datasets.

File	The purpose of the file
Enrolments	This file provides basic information regarding the enrolled learners and staff. It also includes demographic information of learners derived from the learners' responses to the more-about-you survey such as gender, country, age range, highest education level, employment status, and employment area.
Step Activity	This file stores information regarding step activity from learners in the course, e.g. the time when a step is first visited, and the last time a step is marked as completed.
Comments	Information about learners' contributions to the discussion section in each step is stored in this file. It includes the text of the comment and the timestamp corresponding to when the comment was made. It also stores the number of likes associated with a comment.
Question Response	This file holds information about the quiz activity of learners. It stores learners' responses, its correctness and the timestamp associated when answering a quiz.
Team Members	Information about organization staff such as their ids and names are stored in this file.
Peer Review Assignment	This file provides information regarding peer review assignments including when the assignment was first viewed, when it was submitted and the number of reviews associated with the assignment.
Peer Review Reviews	This file provides information about the reviewers on an assignment, including when the review was submitted, the reviewer's id and feedback text on each of the assignment guidelines.
Campaigns	Information about the referral used to advertise a course is stored in this file, following the number of enrolments and active learners for each referral.

Methodology

The main challenge to develop a sustainable, dynamic and near real-time dashboard outside the FutureLearn platform, are the choice of the appropriate automations to keep the data up-to-date for the visualizations and the construction of re-usable building blocks to answer common questions. A novel architecture was developed around a two-stage model: 1) data extraction and pre-processing and 2) dashboard development (see Figure 2). The choice of appropriate tools to enable these stages was driven by previous experience with the creation of dashboards for the Coursera platform combined with a preference for experimenting with different products. Python and R were chosen as the scripting languages. Data was hosted in the cloud (Amazon AWS), consisting of a MySQL database, an R server, a Shiny server and an Apache server to serve content.

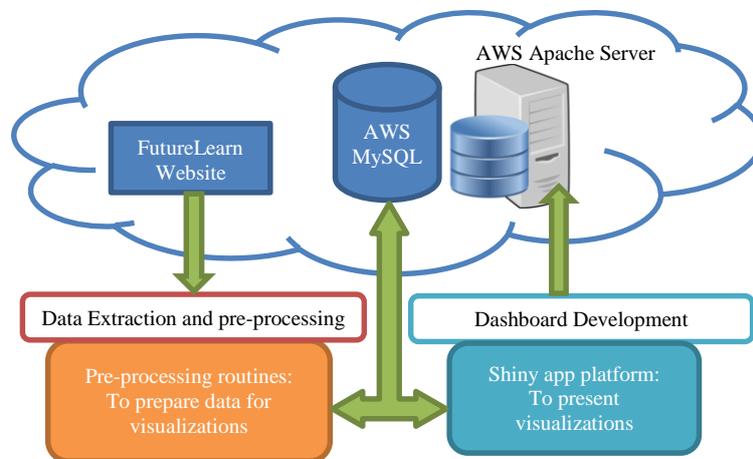
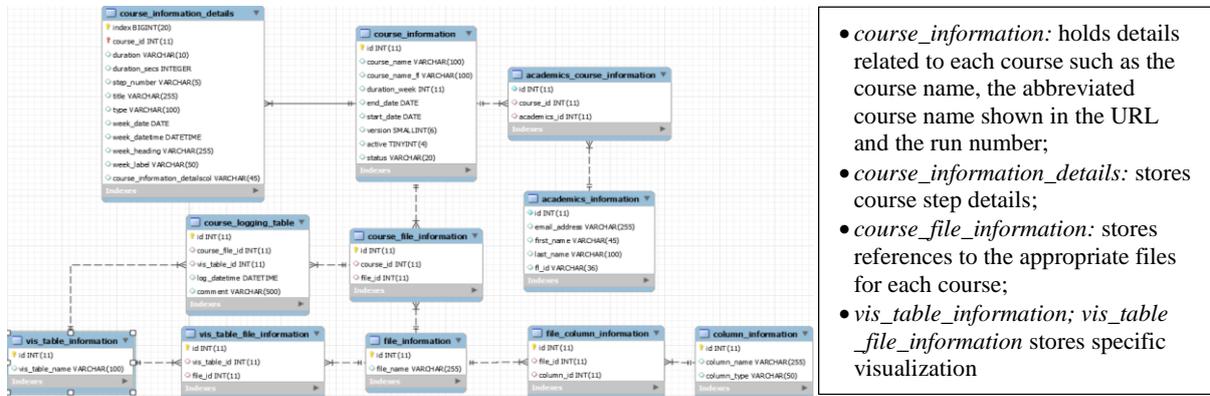


Figure 2. Dashboard architecture for automating dashboard development for FutureLearn MOOCs.

Data Extraction and Pre-Processing

In the data extraction and pre-processing phase, Python scripts were used to automatically login into the FutureLearn platform and download all available files for each course. Each time new CSV files are downloaded, they are stored in a folder, named as per the course title, with previous CSV files archived. To support the automation process for the extraction and processing of raw data, a global database (*futurelearn_courses_information*) was designed, based on a third normal form (3NF) design.

After downloading all course files and loading the data into the database, the preprocessing routines (written as R scripts) prepare the data for the web application and transform the source tables into smaller views, used for individual representations. These are intended to make web data requests more responsive at run-time.



- *course_information*: holds details related to each course such as the course name, the abbreviated course name shown in the URL and the run number;
- *course_information_details*: stores course step details;
- *course_file_information*: stores references to the appropriate files for each course;
- *vis_table_information*; *vis_table_file_information* stores specific visualization

Figure 3: The model of the global database. The core table *course_information* stores metadata on each MOOC. The other tables store metadata regarding the CSV files and the visualization tables.

Exploring different types of visualizations

Previous research (Clayphan et al., 2015; Kia, Pardos, & Hatala, 2016; Stephens-Martinez, Hearst, & Fox, 2014) identified five key areas which appear to direct the attention of educators and developers in MOOCs: 1) An overview of the course, 2) Who are the participants/learners, 3) How participants interact/engage with the material, 4) How participants interact in the forums, and 5) How participants perform in the course.

In the design of our dashboard, we took these areas into account as well as principles for designing dashboards in order to showcase critical information to achieve one or more objectives (Few, 2006). Together with more traditional bar charts, line charts, scattergrams and heatmaps, a few examples are presented below to illustrate the types and modes of representation used. The first example (Figure 4 – left) shows a Sankey chart representing a typical sequence step progression in the course. This is modeled similarly to the representation used in Google analytics. However, in the educational context there is a lack of an appropriate conversion metric (i.e. shopping cart or checkout), therefore the visualization in this context is based on the selection of all pathways of N steps available in the data and allows the end user to select the number of steps. Information about step types are included to show the frequency of transitions from one type to another.

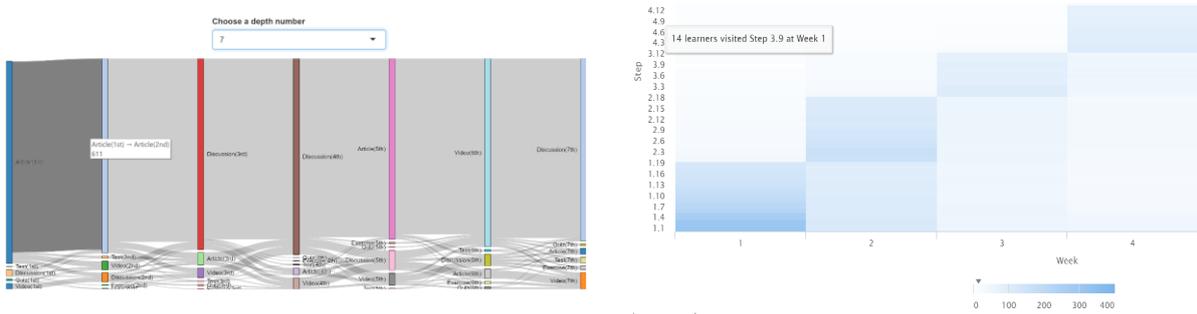


Figure 4. Left: Sankey chart of transitions from step to step labeled by type, Right: patterns of completion by steps (vertical) and weeks (horizontal).

Another interesting representation focuses on learners’ sequences of actions (Figure 4 – right). From the heatmap shown, we see that the typical learner follows the course in a linear fashion, based on the sequence of steps presented (the middle diagonal). Figure 5 shows an alternative representation of transitions between steps. This highlights that learners follow a linear sequence in a course in most steps, but towards the end of a week, there is a high chance of exploratory behavior, as evidenced by jumping to other steps in different weeks.

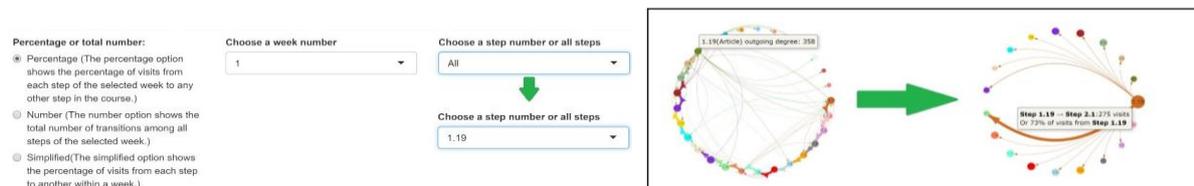


Figure 5. Step transition of learners from each step in the course to another.

The next example (Figure 6 – left) shows the number of comments by date labeled according to step types: this shows that social interaction occurring in ‘discussion’ steps is not the predominant forum for interaction between learners. Additionally, a visualization of sentiment for comments in steps (Figure 6 – right) shows positive and negative scores using a novel algorithm to identify the valence of the comments compared to a reference dictionary and assigning a score to the whole comment, before grouping the comments in the step.



Figure 6: Left: The number of comments on different dates grouped by step types. Right: the sentiment analysis of comments on each step shown as a percentage.

Conclusion and Future Work

This paper demonstrated the rationale and process for beginning to develop an analytics dashboard for the FutureLearn platform delving into some technical details of the implementation. Similar to Leon Urrutia et al., (2016) we opted to develop our visualizations in R, served via a Shiny server. Unlike them however, we focused on creating individual building blocks rather than a complex dashboard interface. This is informed by an attempt to limit the amount of information presented to end-users, a focus on reusability and from our previous experience with end-users from the implementation of dashboards for Coursera MOOCs (Clayphan et al., 2015).

The development process allowed us to explore alternative ways to implement a dashboard that other FutureLearn partner institutions may find useful, with a number of different visualizations explored, as well as consideration of related literature (Stephens-Martinez et al., 2014). Furthermore, a direct comparison with a similar effort by (Leon Urrutia et al., 2016) demonstrates the viability and effectiveness of the implementation. Our implementation provides an opportunity to consider the best ways to use the tool with both educators and students. In the future, opportunities can be easily provided to investigate other sophisticated visualizations and, importantly, test their effectiveness with stakeholders as an activity unfolds, than just at the end of the course.

References

- Bayne, S., & Ross, J. (2014). The pedagogy of the Massive Open Online Course: the UK view. *The Higher Education Academy*, 1-76.
- Clayphan, A., Vigentini, L., Zhang, L., & Zhao, C. (2015). Building Dashboards for Massive Open Online Courses. Paper presented at the Australian Learning Analytics Summer Institute (ALASI), Sydney.
- Cobos, R., Gil, S., Lareo, A., & Vargas, F. A. (2016). Open-DLAs: An Open Dashboard for Learning Analytics. Paper presented at the Proceedings of the Third (2016) ACM Conference on Learning @ Scale, Edinburgh, Scotland, UK.
- Duval, E. (2011). Attention please!: learning analytics for visualization and recommendation. Paper presented at the Proceedings of the 1st International Conference on Learning Analytics and Knowledge, Banff, Alberta, Canada.
- Few, S. (2006). *Information Dashboard Design: The Effective Visual Communication of Data*: O'Reilly Media, Inc.
- Fredericks, C., Lopez, G., Shnyder, V., Rayyan, S., & Seaton, D. (2016). Instructor Dashboards In EdX. Paper presented at the Proceedings of the Third (2016) ACM Conference on Learning @ Scale, Edinburgh, Scotland, UK.
- Kandel, S., Paepcke, A., Hellerstein, J., & Heer, J. (2011). Wrangler: interactive visual specification of data transformation scripts. Paper presented at the Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Vancouver, BC, Canada.
- Kia, F. S., Pardos, Z. A., & Hatala, M. (2016). Learning Dashboard: Bringing Student Background and Performance Online. Paper presented at the Data Literacy for Learning Analytics (DLitLA 2016), Edinburgh, UK.
- Linden, A., & Fenn, J. (2003). Understanding Gartner's hype cycles. *Strategic Analysis Report N° R-20-1971*. Gartner, Inc.

- Leon Urrutia, M., Cobos, R., Dickens, K., White, S., & Davis, H. (2016). Visualising the MOOC experience: a dynamic MOOC dashboard built through institutional collaboration. In *Proceedings of the European MOOC Stakeholder Summit 2016*.
- Macfadyen, L. P., & Dawson, S. (2010). Mining LMS data to develop an “early warning system” for educators: A proof of concept. *Computers & Education*, 54(2), 588-599.
doi:<http://dx.doi.org/10.1016/j.compedu.2009.09.008>
- Ruiz, J. S., #233, D, c. J. P., #237, az, Jos, . . . Kloos, C. D. (2014). Towards the development of a learning analytics extension in open edX. Paper presented at the Proceedings of the Second International Conference on Technological Ecosystems for Enhancing Multiculturality, Salamanca, Spain.
- Shi, C., Fu, S., Chen, Q., & Qu, H. (2014, 25-31 Oct. 2014). VisMOOC: Visualizing video clickstream data from massive open online courses. Paper presented at the Visual Analytics Science and Technology (VAST), 2014 IEEE Conference on.
- Stephens-Martinez, K., Hearst, M. A., & Fox, A. (2014). Monitoring MOOCs: which information sources do instructors value? Paper presented at the Proceedings of the first ACM conference on Learning @ scale conference, Atlanta, Georgia, USA.
- Verbert, K., Govaerts, S., Duval, E., Santos, J. L., Van Assche, F., Parra, G., & Klerkx, J. (2014). Learning dashboards: an overview and future research opportunities. *Personal and Ubiquitous Computing*, 18(6), 1499-1514. doi:10.1007/s00779-013-0751-2
- Yuan, L., & Powell, S. (2013). MOOCs and open education: Implications for higher education. Retrieved on 1/9/2016 from <http://publications.cetis.ac.uk/wp-content/uploads/2013/03/MOOCs-and-Open-Education.pdf>

Please cite as: Chitsaz, M., Vigentini, L. & Clayphan, A. (2016). Toward the development of a dynamic dashboard for FutureLearn MOOCs: insights and directions. In S. Barker, S. Dawson, A. Pardo, & C. Colvin (Eds.), *Show Me The Learning. Proceedings ASCILITE 2016 Adelaide* (pp. 116-121).

Note: All published papers are refereed, having undergone a double-blind peer-review process.



The author(s) assign a Creative Commons by attribution licence enabling others to distribute, remix, tweak, and build upon their work, even commercially, as long as credit is given to the author(s) for the original creation.