So, H. J. et al. (Eds.) (2020). Proceedings of the 28th International Conference on Computers in Education. Asia-Pacific Society for Computers in Education

Reimagining the conceptualisation, design and delivery of Learning Analytics

Aneesha BAKHARIA[a*] & Linda CORRIN[b]
[a]*The University of Queensland, Brisbane, Australia*
[b]*Swinburne University of Technology, Melbourne, Australia*
*aneesha.bakharia@gmail.com

Abstract: Learning analytics applications have the potential to bring value to educational environments by providing the ability to analyse and present data about students in ways that are meaningful and actionable for educators. Recent studies have highlighted the importance of involving educators in the design process for these systems to ensure that what is developed meets their needs. However, this can add significant time and resource investment to a development project and requires the involvement of a range of people with data science and developer skills. Across various industries, new approaches are being imagined and trialled to reduce reliance on technical staff and long development processes when designing data-based systems instead enabling users to create a story from data without needing programming skills. In this paper we review the innovations introduced by DIVE, a mixed initiative visualisation and analysis application developed by a team at the MIT Media Lab, and discuss how these can be adapted to the field of learning analytics. DIVE includes a range of features that can provide value to learning analytics applications such as automated dataset statistical analysis, visualisation recommendation, and data story generation. An analysis of these features informs a set of guidelines which will enable the reconceptualisation of co-design sessions, reducing the time required to design learning analytics applications and facilitating the automated generation of prototypes, and in some cases full dashboards and analytic applications. The paper concludes with a discussion of future enhancements that could be made to the functionality of DIVE, such as the need for a semantic layer on trace and assessment data, to enable the translation of data into meaningful insights on learning.

Keywords: Learning Analytics, Dashboards, Visualisation Recommendation, Data Stories

1. Introduction

It takes time to design and mainstream a successful learning analytics application! The time investment is further extended when adopting an iterative co-design process, an approach which is rapidly growing in popularity in the learning analytics literature (e.g., Dollinger et. al., 2019, Shobani et. al., 2019). Co-design with educators ensures that the application directly addresses their requirements, resulting in an end product that provides more meaningful and actionable outputs to enable educators to support student learning. The increase in the use of co-design methods has also led to the development of tools to support the resulting design discussions. One example is a card-based design approach (Alvarez, et. al., 2020) where educators work through a series of design steps using cards relating to each area of the proposed application to identify useful components. While such approaches are able to trigger relevant discussion, they are only able to facilitate the abstract assembly of ideas. Much work still remains to develop the application and requires a diverse skill set (e.g., data science and developer skills) after the co-design session has ended to prototype, review and implement the application.

A further challenge in the design of learning analytics applications is enabling educators to make meaning from the analysis of different types of data (e.g. clickstream event data, assessment data, etc.) in a way that relates the outcomes of analyses of the data to what we know about how students learn. Within the learning analytics community this is known as the ‘clicks to construct’ issue (Knight & Shum, 2017). In order to be able to translate the data and analyses into meaningful insight there is an important role that learning design and theory plays in providing context and an epistemological frame for interpretation (Knight, Shum & Littleton, 2014). Bringing together the data and the theory enables learning analytics practitioners and educators to build stories from the data that aid in understanding students' actions and identifying appropriate interventions that can be made (Echeverria et al., 2018). However, this is a complex and sophisticated process which is especially hard to implement in systems that are designed to address a range of educator queries and needs.

While the complexity of learning is an ongoing challenge within the field of learning analytics, the other challenges identified here around the application development process are also present across a number of industries. As a result, the data science community continues to look for innovative ways to address these challenges, in particular to lower the technical and time barriers to allow users to easily build stories from data. As a field that grew from a multi-disciplinary foundation, learning analytics practitioners and researchers often look for innovations across disciplines and industries to inspire new ways of analysing and understanding education. In this paper we review recent innovations that have been made within data science research to improve the learning analytics application development workflow.

The review in this paper relates to three fundamental questions that are essential to re-conceptualising the design and delivery of learning analytics applications:

1. Is it possible to produce working prototypes within a design session between a learning analytics practitioner and educators?
2. Can learning analytic applications be automatically generated?
3. Can the generated learning analytics application address the 'clicks to construct' issue?

In order to answer these questions we first review a recently proposed application called DIVE which includes a workflow that would be suitable for the design of learning analytics applications. We then suggest improvements, and propose a set of detailed guidelines for adapting DIVE to a learning analytics setting and finally outline two proposed workflows within a higher education setting.

2. An Overview of DIVE

DIVE is a web-based platform that supports data ingestion, visualisation, analysis and storytelling. It was developed by Kevin Zeng Hu and César A. Hidalgo at the MIT Media Lab (https://www.media.mit.edu/projects/dive/overview/) to allow users to explore and use data to create stories without having to know how to code. DIVE provides a mixed-initiative workflow for analysts, including modern techniques such as visualisation recommendation and field level data model inference. In the DIVE evaluation study analysts were given a set of predefined visualisations and analytics to perform in DIVE and Microsoft Excel, with DIVE users showing a significant reduction in time to complete the tasks (Hu, et. al., 2018). DIVE has been selected to be reviewed as the data exploration workflow can be adapted to learning analytics and also because both the client and server-side code are open source. DIVE is implemented in Flask and React.

Figure 1 shows the DIVE user interface. The left bar (A) provides an intuitive way to navigate through the workflow provided by the application. The top bar (B) displays the current project and dataset being analysed. The main screen (C) displays the content of the selected workflow element with configurable parameters available on the right hand bar (D). The various sections of the DIVE user interface remain consistent across all supported workflows. Figure 2, illustrates the various workflows supported by DIVE.

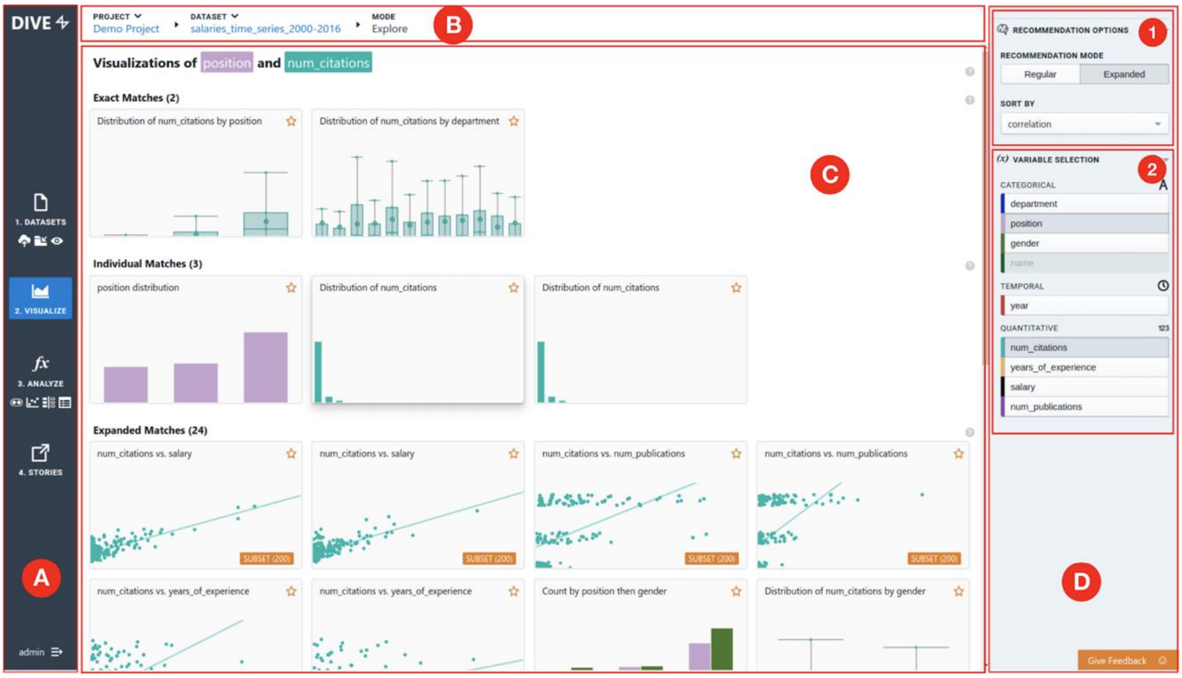

Figure 1. The DIVE user interface, reproduced from (Hu, et. al., 2018).

2.1 Data Ingestion and Inspection

DIVE is able to import tabular data, perform data model inference and provide statistical analysis of each column. It categorises columns using a heuristic-based algorithm and uses this as additional information within any automated analysis. The basic categories are nominal, ordinal, and continuous along with a semantic categorisation which includes categorical, temporal, and quantitative. After data is imported, the basic statistical properties of each column are shown to the user. For categorical fields the number of unique values and the frequency of each value is displayed. Summary statistics are shown with simple histogram visualisations for ordinal and continuous fields.

2.2 Visualisation and Drilldown

DIVE provides mixed-initiative visualisation recommendations. The user specifies fields (or columns) of interest and the DIVE visualisation recommendation engine produces a set of visualisations that may be of interest to the user ranked by statistical measures. If no fields are selected, univariate descriptive visualisations are produced for all fields in the uploaded dataset. DIVE automatically generates charts which can be stored and reused within a story. There is no requirement for the user to know how to bind or encode data for display on a chart or have knowledge of using a charting library.

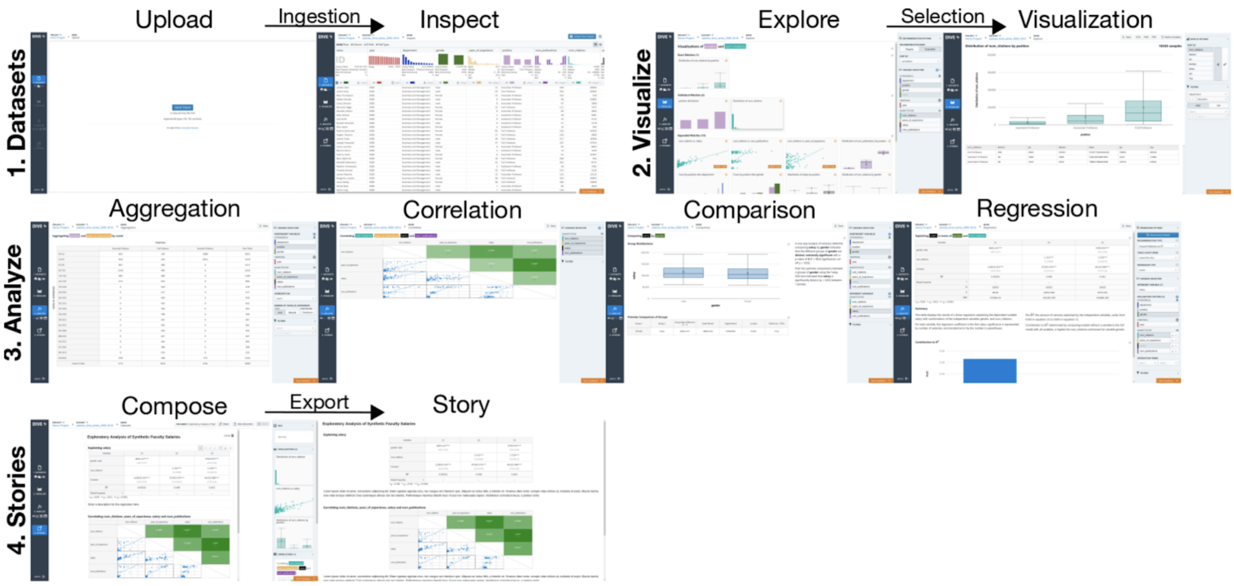

Figure 2: Workflows supported in DIVE, reproduced from (Hu, et. al., 2018).

2.3 Analysis

DIVE is able to perform common statistical functions on the uploaded tabular data. The main functionality included is the ability to aggregate data (i.e., calculate mean and standard deviation), view the correlation between two variables, compare sub-groups using one-way or two-way ANOVA, and perform linear or logistic regression.

2.4 Story Creation and Sharing

Any visualisation or analysis result can be saved and assembled into a story. A story can then be shared via a public link. The inclusion of compose functionality makes DIVE an authoring and publishing tool. DIVE allows users to easily analyse data and share their findings, though no security (i.e., authorisation and authentication) is enforced for shared stories.

3. Extending the Concepts Introduced in DIVE in Learning Analytics

The three main innovations introduced in DIVE include data model inference, mixed-initiative visualisation recommendation, and the generation of sharable analytic narratives. All three are very relevant to learning analytics. The automated data model inference and mixed-initiative visualisation recommendation in particular would allow prototype dashboards and applications to be built in co-design sessions. Educators participating in co-design sessions would be able to select or suggest the variables of interest and the design facilitator or even the educator, or a support member of the teaching team could see a preview in real time. This nicely addresses the challenges of time and technical skill in progressing through an application design process. There are however opportunities to further enhance DIVE to better suit the learning analytics domain.

3.1 Data Ingestion and Inspection

DIVE uses a heuristic algorithm to determine the semantic categorization of columns (i.e., fields) in a tabular dataset. The current algorithm would need to be tested with clickstream data (originating from learning management systems and embedded tools such as Echo360), student demographic, enrolment and assessment data. DIVE associates a single tabular dataset with a project. The data model would need to be extended for learning analytics to allow multiple related datasets. This would allow for datasets from different systems to be uploaded and linked via a student identifier.

Within one-on-one sessions with an educator or multiple participant co-design sessions, real student datasets may not be available or unable to be used due to privacy concerns. In order for the tool to be useful within either setting, functionality is required to be able to synthesise data for various related student datasets. It is envisioned that the fields for dummy datasets would need to be entered along with the parameter configurations (e.g., max, min and standard deviation). An additional requirement would be to allow for student sub-groups to be modelled as this would allow for discussion around the interpretation of visualisations and provision of personalised feedback within co-design sessions. Data synthesis is both under researched and utilised within learning analytics research but ideas from recent open source data synthesize tools such as Synner (Mannino & Abouzied, 2020) can be incorporated within DIVE. Synner provides a declarative user-interface for users to describe the statistical distribution of fields and the relationship between fields.

3.2 Visualisation and Drilldown

In recent years many visualisation recommendation algorithms have been proposed (Wongsuphasawat et. al., 2020, Lima et. al., 2020) and the algorithm within DIVE can be updated to incorporate ideas from recent research. A key feature of the visualisation engine embedded within DIVE is that mixed-initiative recommendations are supported. The mixed-initiative feature is important because unlike other recommendation engines that automatically select the variables to be displayed in a visualisation, DIVE allows the user to specify the fields they are interested in. Within a co-design session, this feature will facilitate the inclusion of feedback from the educator thereby directing the visualisations to be generated.

3.3 Story Creation and Sharing

The Compose feature in DIVE allows visualisations to be collated and annotated on a web page. While this is a simple idea, it essentially completes the full workflow from design to a deployed analytics artefact. Imagine the power of generating a custom course-specific dashboard for an educator in a co-design session and creating a working analytics tool that can be re-used with real data for the forthcoming semester. However, an essential prerequisite for implementing DIVE within a university setting is password protection of shared analytics which would need to involve integration with university single sign on to address privacy issues relating to student demographic and assessment data.

4. Addressing the “Clicks to Construct” Issue in Learning Analytics

The automated recommendation of time series charts however will not necessarily provide the insight that educators need to adapt courses and provide personalised feedback. While trace data (also known as clickstream data) has proven difficult to map to higher level constructs (i.e., the “clicks to construct” issue) (Knight and Shum, 2017), recent learning analytics research has focussed on addressing the issue by adding linked data or metadata to the events captured within a clickstream.

In this paper, we use the term “semantic layer” to describe the inclusion of linked or metadata to develop meaningful constructs for visualisations and metrics. DIVE’s data model would need to be extended to allow for additional linked and metadata to be uploaded or synthesised (i.e., dummy data). It is envisaged that initially three semantic layers that use techniques proposed from recent learning analytics research (i.e., 2019) could be added to the DIVE data model. Two (2) of the techniques require additional metadata (e.g., educator annotations of course material to scheduled delivery week and estimates of activity duration) while the additional one (1) proposes an alternate way of processing student assessment data to generate a sequential Sankey diagram.

4.1 *Classifying Student Activity as Ahead, Preparing, Catching Up and Revising*

In Gašević et, al., (2019), a novel algorithm to classify student interaction within blended or fully online course material is proposed. Clickstream data is enriched by linking events to content modules and their scheduled delivery date. The additional metadata allows student access to be classified as being ahead, preparing, catching up, revisiting or not accessing the course (introduced to provide a 100% student population representation). The classification can be visualised (see Figure 3) to provide a holistic weekly overview of a student cohorts progress with coursework. A Sankey visualisation can also be used to display student transitions through semester weeks.

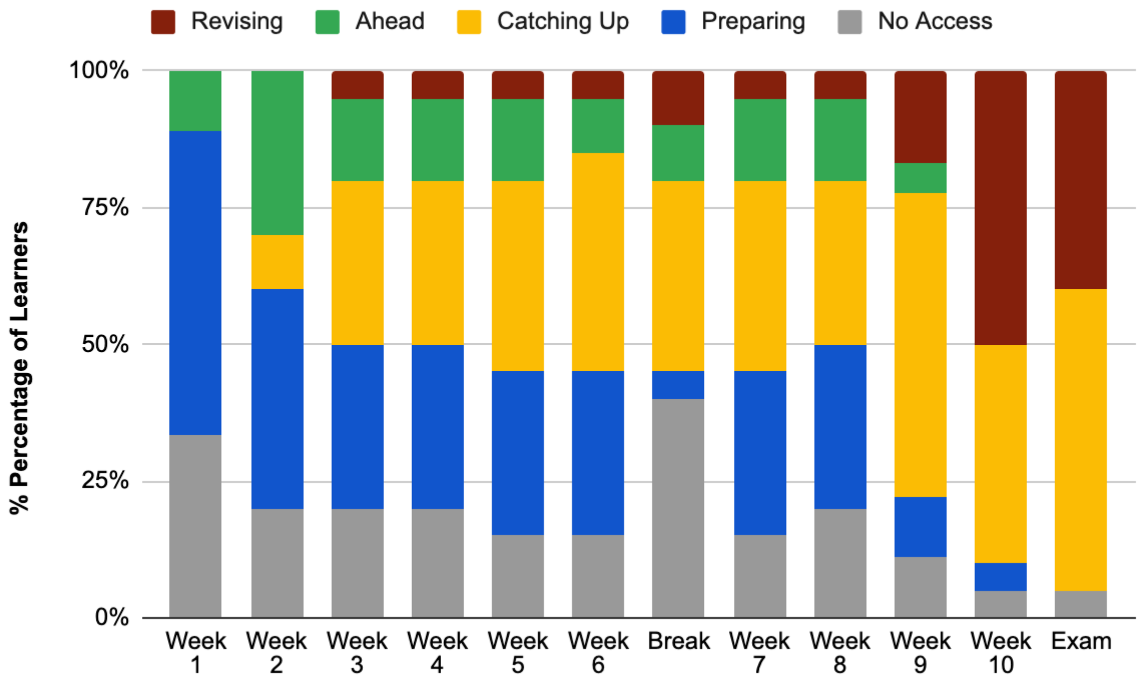

Figure 3: Holistic overview of progress made by cohort of students in completing online content.

4.2 *Comparing Activity Time Estimates to Actual Student Completion Time*

Ginda et, al., (2019) enriches data relating to course access events with the course section being completed, the average time it has taken students to complete, and an estimate of activity completion time provided by the course designer. The resulting visualization (see Figure 4), allows educators to easily identify sections and activities that are either completed too quickly or slowly thereby increasing the estimated time for students to complete the module.

4.3 *Analyse Student Grade Transitions*

The grade Sankey visualization introduced by Deng et, al., (2019) provides new insight on progression and performance in course assessment. The grade Sankey is created by calculating how students transition between grades for each assessment item and then displaying the transition flow on a sequential Sankey diagram with fixed axis labels for each assessment item. The visualization provides a way to identify and provide feedback and support to students with decreasing grades or notable changes in performance.

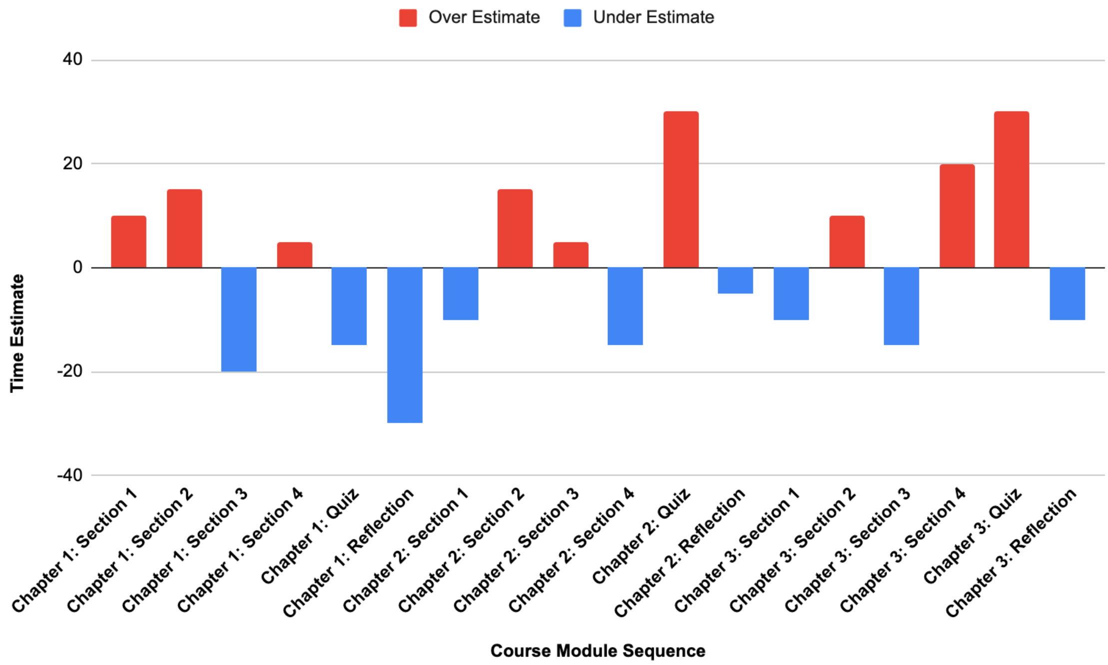

Figure 4: Difference between activity time estimates and actual time spent by students.

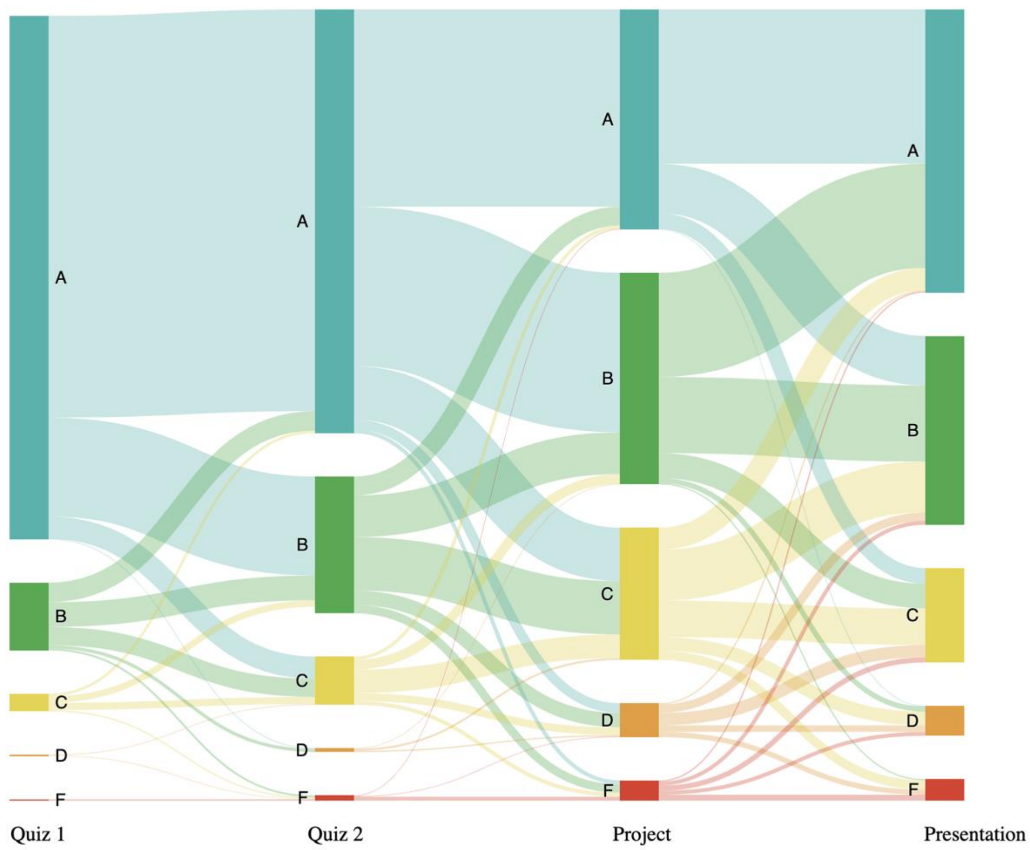

Figure 5: Sequential Sankey used to illustrate student grade transitions.

5. Use Cases within Learning Analytics

In this section we present two (2) example use cases for using DIVE after enhancements for learning analytics have been implemented within the tool. The examples serve to showcase the potential of the key innovations introduced in DIVE and demonstrate that the proposed workflow is flexible enough to be used in a variety of learning analytics design sessions.

5.1 *Educational Data Scientist Works with Teaching Team in a Co-Design Session*

A teaching team consisting of a course coordinator, lecturers teaching sections of the course and the faculty assigned learning designer has just completed restructuring a course for flipped delivery (i.e. online modules that need to be completed prior to active face to face workshops). The course coordinator, nervous about delivering the course for the first time, contacts the central learning analytics unit at the university. The university while investing in commercial analytics products has also placed an emphasis on developing custom co-designed solutions to better meet the analytics needs of specific domains. The central learning analytics unit proposes a co-design session with the faculty assigned educational data scientist.

The co-design session begins with the educational data scientist asking questions about the course redesign. No real student data has been collected as the course has not yet run, but the educational data scientist is able to synthesize mock (i.e., dummy) data for each assessment item and the student event clickstream for each of the online modules within DIVE. The educational data scientist then clicks on the visualisations tab and asks the team what questions they would like to answer. The lecturers would like to know whether students are completing the online module before they come to campus so that they can adapt the weekly face to face workshops accordingly. The educational data scientist then uses the mixed-initiative visualisation recommendation engine by selecting the variables of interest in the clickstream and applies a semantic layer that includes information on whether students have completed the online module in the scheduled delivery week. The recommended visualisations include a time series plot of access, a bar chart of the % of students that accessed content per teaching week, a box plot chart showing the percentage of students that have completed each module, and a stacked bar chart showing the percentage of students that were categorised each week according to whether they were on track, ahead, catching up or revising content. The teaching team all agree that the stacked bar charts and the box plot both provide a good overview of the whole cohort. The educational data scientist saves the stacked bar chart and the box plot and also informs the team that in order to get these meaningful charts they will need to map each module from the LMS to a teaching week. The educational data scientist repeats the process of asking questions, specifying variables of interest and saving recommended visualisations as the co-design session progresses.

At the conclusion of the co-design session, all of the visualisations are added to a dashboard that is then shared with the team. The educational designer informs the team that the dashboard will automatically be populated with real student data once the semester begins.

5.2 Educator Created Narrative Visualisation

An educator would like to make a case for redesigning the assessments and their weightings in a course. The educator has a sense that students who get a good grade in early assessment items are performing badly in the final exam. The educator has a download of the Grade Centre file from the Blackboard LMS. Viewing the data in Excel does not directly reveal any insights. The educator is aware of a new tool called DIVE being piloted at the university. The educator logs into the tool and uploads the gradebook file. DIVE is able to automatically detect assessment scores and matches these to a semantic layer that can calculate the grade transitions of each student. On the visualisation recommendation page, the various histograms are proposed including histograms of scores for each assessment, boxplots for each assessment on a single chart, and a sequential Sankey diagram. The educator thinks that the sequential Sankey is able to illustrate clearly that the final assessment task is either too difficult or incorrectly weighted and proceeds to add the visualisation to a story. The educator annotates the visualisation and saves a version for inclusion in a presentation for the course review panel.

6. Discussion

After reviewing DIVE, suggesting enhancements specifically for learning analytics and addressing the ‘clicks to construct’ issue in learning analytics, we are now able to address the key research questions.

1. Is it possible to produce working prototypes within a design session between a learning analytics practitioner and educators?
   From our review of the functionality of DIVE it would seem that the platform could facilitate the development of working prototypes for educators, however enhancements would be needed to make the platform more suitable to the context of learning analytics. The workflow proposed by DIVE (i.e., Ingest and Inspect, Visualise, Analyse and Compose) simplifies data exploration and data story authoring without requiring specialist skill sets (i.e. data science or programming skills). Key enhancements are however required to contextualise DIVE for use in learning analytic co-design sessions including data model inference for clickstream and assessment data and the ability to synthesise dummy data. Mixed-initiative visualisation recommendation was also preferable instead of a simple gallery of visualisations, where the variables of interest could be specified, which would allow co-design session participants to direct the recommended visualisations.
2. Can learning analytic applications be automatically generated?
   The key innovations that facilitated the automatic generation of visualisations within DIVE were the data inference model and the visualisation recommendation engine. The data inference model was able to classify fields within the uploaded dataset which was then used by the visualisation recommendation engine. The visualisation recommendation engine used the field level classifications to generate appropriate variable encodings for representation on charts and visualisations. Incorporating the ability to generate charts directly from uploaded data is essentially what provides the functionality to automatically generate an application without the underlying programming knowledge. Visualisation recommendation is an active research area and requires further investigation by the learning analytics research community.
3. Can the generated learning analytics application address the ‘clicks to construct’ issue?
   The current functionality and workflow provided by DIVE would be directly applicable to most domain areas, however some enhancements would be required to address this issue more specifically. One of the challenges with learning analytics is that trace data of discrete student events with a learning system provides very little insights into the actual learning of students. In order to address this issue latest research on higher level constructs needs to be incorporated into the data model inference and aggregate analysis modules in DIVE. Higher level constructs are usually created by including a semantic layer mapping discrete clickstream events to additional metadata. Within this paper we proposed an initial set of three (3) layers, with the aim of increasing this list as new research emerges.

7. Conclusion and Future Directions

The DIVE application has served as a good example of how a modern data exploration workflow tool can be reviewed and adapted to support learning analytics workflows. The innovations introduced within DIVE include data model inference, mixed-initiative visualisation recommendation and the ability to assemble visualisation into shareable narrative stories. These innovations coupled with the adaptations proposed in this paper, have the ability to transform the current design, development and deployment processes currently used in learning analytics.

Future research involves implementing the enhancements within DIVE and conducting a series of case studies with practitioners, educators, learning designers, educational data scientists and data engineers to evaluate the processes discussed here in practice. It is hoped that there will also be interest from the learning analytics research community to further explore research in visualisation recommendation and data story creation.

References

Alvarez, C. P., Martinez-Maldonado, R., & Shum, S. B. (2020, March). LA-DECK: A card-based learning analytics co-design tool. In *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge* (pp. 63-72).

Deng, H., Wang, X., Guo, Z., Decker, A., Duan, X., Wang, C., ... & Abbott, K. (2019). PerformanceVis: Visual analytics of student performance data from an introductory chemistry course. *Visual Informatics*, *3*(4), 166-176.

Dollinger, M., Liu, D., Arthars, N., & Lodge, J. (2019). Working Together in Learning Analytics Towards the Co-Creation of Value. *Journal of Learning Analytics*, *6*(2), 10–26.

Echeverria, V., Martinez-Maldonado, R., Granda, R., Chiluiza, K., Conati, C., & Shum, S. B. (2018, March). Driving data storytelling from learning design. In *Proceedings of the 8th international conference on learning analytics and knowledge* (pp. 131-140).

Gašević, D., Matcha, W., Jovanović, J., & Pardo, A. (2019). Analytics of time management strategies in a flipped classroom. *Journal of Computer Assisted Learning*.

Ginda, M., Richey, M. C., Cousino, M., & Börner, K. (2019). Visualizing learner engagement, performance, and trajectories to evaluate and optimize online course design. *PloS one*, *14*(5), e0215964-e0215964.

Hu, K., Orghian, D., & Hidalgo, C. (2018, June). Dive: A mixed-initiative system supporting integrated data exploration workflows. In *Proceedings of the Workshop on Human-In-the-Loop Data Analytics* (pp. 1-7).

Knight, S., & Shum, S. B. (2017). Theory and learning analytics. *Handbook of learning analytics*, 17-22.

Knight, S., Shum, S. B., & Littleton, K. (2014). Epistemology, assessment, pedagogy: where learning meets analytics in the middle space. *Journal of Learning Analytics*, *1*(2), 23-47.

Mannino, M., & Abouzied, A. (2020, June). Synner: Generating Realistic Synthetic Data. In *Proceedings of the 2020 ACM SIGMOD International Conference on Management of Data* (pp. 2749-2752).

Jovanovic, J., Mirriahi, N., Gašević, D., Dawson, S., & Pardo, A. (2019). Predictive power of regularity of pre-class activities in a flipped classroom. *Computers & Education*, *134*, 156-168.

Lima, R. D. A., & Barbosa, S. D. J. (2020). VisMaker: A Question-Oriented Visualization Recommender System for Data Exploration. *arXiv preprint arXiv:2002.06125*.

Shibani, A., Knight, S., & Shum, S. B. (2020). Educator perspectives on learning analytics in classroom practice. *The Internet and Higher Education*, *46*, 100730.

Wongsuphasawat, K., Moritz, D., Anand, A., Mackinlay, J., Howe, B., & Heer, J. (2015). Voyager: Exploratory analysis via faceted browsing of visualization recommendations. *IEEE transactions on visualization and computer graphics*, *22*(1), 649-658.