A Web-based Data Visualization Tool Regarding School Dropouts and User Assessment

Angelika Mark Kayanda
The Nelson Mandela African Institution of Science and Technology
Arusha, Tanzania
kayandaa@nm-aist.ac.tz

Dina Machuve
The Nelson Mandela African Institution of Science and Technology
Arusha, Tanzania
dina.machuve@nm-aist.ac.tz

Abstract—Data visualization is important for understanding the enormous amount of data generated daily. The education domain generates and owns huge amounts of data. Presentation of these data in a way that gives users quick and meaningful insights is very important. One of the biggest challenges in education is school dropouts, which is observed from basic education levels to colleges and universities. This paper presents a web-based data visualization tool for school dropouts in Tanzania targeting primary and secondary schools, together with the users’ feedback regarding the developed tool. We collected data from the United Republic of Tanzania Government Open Data Portal and the President’s Office - Regional Administration and Local Government (PO-RALG). Python was then used to preprocess the data, and finally, with JavaScript, a web-based tool was developed for data visualization. User acceptance testing was conducted and the majority agreed that data visualization is very helpful for quickly understanding data, reporting, and decision making. It was also noted that the developed tool could be useful not only in the education domain but it could also be adopted by other departments and organizations of the government.

Keywords—data visualization; school dropouts; primary schools; secondary schools

I. INTRODUCTION

Dropping out from school is a global education issue. Student retention and completion remain a challenge in education sector, causing a significant loss in current and future intellect. The United Republic of Tanzania through the Government Open Data Portal estimated 56,573 primary school dropouts in the year 2013 and a significant increase to 85,985 in 2015 and 117,927 in 2016 [1]. Secondary school dropouts were about 94,990 in 2012, 61,488 in 2015 and 65,700 in 2017 [2]. The numbers remain high even after Tanzania started implementing the free basic education policy in 2016. Though studies about school dropouts have been done, and data have been collected and published, the data have not yet been fully explored. A study on school dropouts in three regions of Tanzania revealed that socio-economic and political factors, along with parents’ views and government contribute to school dropouts [3]. The study results were presented in percentages, tables and simple text. Similarly, a study focused on the artisanal mining areas of the Geita region in Tanzania presented the data regarding the dropout causes in mostly text

[4]. This kind of presentation provides limited interactivity and difficulty in interpreting the information or revealing trends which are helpful in decision making. A recent study used machine learning techniques to develop a predictive model for secondary school dropouts in Tanzania to help identify students who are at risk of dropping out and embedded a visualization module. However, the module visualized school data from only five regions in Tanzania [5].

Recently, data visualization has gained significant attention, not only for its ability to draw meaningful insights from data and events occurring around us through visual displays [6], but also for its ability to take advantage of the human mind’s powerful visual processing capabilities [7, 8]. There are many works that underline the importance of data visualization showing its usefulness and providing insights into complex datasets, by communicating key aspects in a more intuitive and meaningful way [9-11]. The United Republic of Tanzania Government Open Data Portal has visualized maps of primary schools’ pupil-teacher ratio, final examination performance, pupil-classroom ratio, and form four examination performance for the year 2014, but it is missing visualizations for dropouts [1, 2]. The Tanzania National Bureau of Statistics has presented the status regarding dropouts of population aged above 5 years, but no visualization can be found [12]. Therefore, the present student dropout data, which have been lightly explored, can be made more useful by employing visualization to support reporting, decision making and ultimately benefit the overall education sector and other interested parties. In this paper, we divided the task into three objectives: data preparation, implementation, and testing. After the data were collected and preprocessed, rapid prototyping was used for the development, and finally feedback was gathered from the users through focus group discussion and questionnaires. Herein we report the experimental results of a data visualization tool developed for school dropouts in Tanzania.

II. RELATED WORK

A. Visualization in Education

Data visualization has been employed in various aspects within the education domain. A visualization tool was developed for Dropout Reasoning and Prediction for the
Massive Open Online Courses (MOOCs) [13]. The Pearson’s Learning Curve provides dynamic, interactive and easy-to-use data visualization tools to enable researchers and policymakers to derive meaning from global education datasets [6]. Furthermore, visualization was useful in unearthing interesting findings and more detailed depiction of learners’ behaviors in an adaptive learning environment, which would otherwise remain hidden [14]. For improved teaching and learning online, two interactive visualization tools were developed at the Unitelma Sapienza University and the IISLab of Tampere University of Technology for the Moodle LMS as plug-ins. These tools assist students and teachers to monitor, evaluate and make decisions to improve the learning outcome [15]. Discrete graphs have been used to visualize learning logs, a method that collected data from e-text and learning management systems. The graphs assisted in observing learning activities’ features for each grade by visualizing combinations of achievements and failures, thus helping discovering the learning activities that should be avoided by students [16]. A web platform named SWATShare, designed to enable collaborative hydrology research and education online, makes use of visualization tools such as dynamic time-series plots and spatial maps, to help users understand the variability of various hydrological processes through time [17]. Furthermore, one study showed that integrating teaching strategies with interactive visualization techniques could greatly benefit the education domain and enhance visualization literacy to children in early grades [18]. It is pointed out that even though visualizations have recently become common, the ability of youths and adults to interpret them is still relatively low, which can be a serious handicap to learning and making informed decisions.

In education, data visualization has been shown to play a key part in identifying and analyzing key causes and their variations in various issues, thus assisting decision makers to enforce more effective schemes in teaching, monitoring learner’s behavior and improve learner’s understanding of the learning outcomes. Moreover, the usefulness and success of interactive visualizations in online learning environments suggests that it could also be useful in non-online educational issues.

B. Visualization in Other Domains

A study was conducted on the prevalence of visualization for electronic health records data from 1996 to 2013. The study found that although electronic health records are increasingly being visualized, not many techniques have been found to display these complex data effectively and efficiently [19]. A visualization plugin that equipped the translational research open data warehouse with dynamic visual analytical workflows by using modern web technologies such as AngularJS and D3 was developed in [20]. According to Haehn, visualizing connectomics which have massive datasets of brain tissue details, allows neuroscientists to focus on the insights provided by connectome data, rather than managing of these massive datasets. These dynamic and interactive data visualizations provide easier access to data and support post-processing exploration for easier hypothesis generation [21]. A web application called NGL Viewer was developed for the visualization of macromolecular structures. Embracing the modern web technologies, the viewer can assist scientists to easily access 3D structural data [22]. A direct-manipulation approach for non-programmers to understand neural networks has been illustrated through an interactive open source visualization called TensorFlow Playground. This visualization intuitively gives its users a hand-on feel of how neural nets work without any coding [23]. In medical physics, a 3D visualization technique named Cinematic Rendering has been used to show photorealistic representations of 3D images from traditional CT and MRI data [24]. In software engineering, tag cloud has been used as a visualization technique to assist software engineers understand and manage software code [25]. In various studies, visualizations have been used to convey study results and their comparisons in a way that is easily understandable [26-28]. All these works bring to light the data visualization tools in other domains and the technologies employed to create them, whereby some of the technologies could also be applied in the current study.

C. Web-based Visualizations in Tanzania

In Tanzania, a number of visualization tools and interactive dashboards can be currently seen in many web-based systems and portals. The National Bureau of Statistics has a data visualization section with interactive visualizations for Tanzania’s statistics on population, economy, health, environment, agriculture, and education. However, for the case of dropouts, only school dropout status of children with an age above 5 years is shown in numbers with no visualization [12]. The Tanzania National Health Portal of the Ministry of Health Community Development Gender Elderly and Children has a data statistics section which contains interactive graphs, charts, and maps for various health statistics [29]. Furthermore, the Government of Tanzania and the World Bank have developed dashboards for three sectors: health, water, and education, aiming to support reporting and decision making through interactive mechanisms. Regarding the case of education, visualizations for school dropouts are missing [1, 2].

Therefore, data visualization has been widely applied in many domains, but in the educational domain the present visualization tools are mostly for e-learning platforms and the MOOCs. Moreover, most visualization tools for education either do not present student dropout data or present only primary school data, or secondary school data or girl dropout data [5, 28, 30]. This study extends the already done works on student dropouts, aiming to develop a visualization tool for presenting both primary and secondary school dropout data. The tool presents dropout data of the whole country at region and district level through maps, and various kind of charts. The tool further allows users to upload their own data and generate visualizations. This tool will help reporting and decision making in student retention issues in Tanzania.

III. METHODS

Different approaches are widely used in software tool development, utilizing techniques such as, but not limited to, parallel development, prototyping and iterative development [31]. This study used the rapid prototyping approach to develop the web based tool, in which an initial prototype was
Kayanda & Machuve: A Web-based Data Visualization Tool Regarding School Dropouts and User Interactions

developed, evaluated and continuously improved through a number of iterations until the desired outcome was reached. Figure 1 shows the stages followed during the development of the visualization tool.

![Development stages of the visualization tool.](image)

For this study, secondary data were collected from the United Republic of Tanzania Basic Statistics Portal and the PO-RALG. The collected datasets were .csv files. During data collection, a short interview was conducted at the Education Department of PO-RALG to understand their procedure of data collection, analysis, visualization, and publishing of student dropout statistics. Python libraries including pandas, matplotlib, seaborn and bokeh were used to preprocess the data. The resulting datasets, after preprocessing, included two datasets compiled according to level (region-level and council-level) and two datasets compiled according to category (primary schools and secondary schools). The selected features from the datasets included male enrollment, female enrollment, total enrollment, male dropouts, female dropouts, total dropouts, pupil-teacher ratio (PTR), pupil-qualified teacher ratio (PQTR), pupil-classroom ratio (PCR) and pupil-latrine ratio (PLR). The overall design of the data flow and the interactions in the tool are presented in the visualization pipeline shown in Figure 2, which is modified from [32, 33].

![Visualization pipeline.](image)

User interactions in the tool are further indicated in the use case diagram in Figure 3. The users of the tool include decision makers in the education domain such as education officers, education statistics experts, heads of schools, along with teachers, parents, students, education researchers and any citizen interested in education information. Through a web browser, the users can access and view the visualizations using various menus provided in the user interface. Users can further upload data and view the resulting visualizations.

![Use case diagram.](image)

During the development, a number of tools and technologies were employed including JavaScript as the development language, React.js for the front-end development and Node.js along with Express.js for handling the back-end. An online csv to json converter was used to change the preprocessed data to JSON format, which were further formatted and saved as JavaScript files inside the project directory. This allowed the data to be imported to the application components easily. Focus group discussion was then used to solicit users’ feedback on the developed web-based visualizations from the participants. Thirteen participants were selected for the focus group based on their role in basic education and education statistics. Questionnaires were used to assess the users’ opinion on the developed visualization tool for the purpose of supplementing on the feedback from the focus group discussion as shown in Table I.

### Table I. Focus Group Discussion Participants and Questionnaire Respondents

<table>
<thead>
<tr>
<th>Title</th>
<th>Focus group participants</th>
<th>Questionnaire respondents</th>
</tr>
</thead>
<tbody>
<tr>
<td>Principal Education Officers (PEO)</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Principal statisticians</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Heads of school</td>
<td>4</td>
<td>28</td>
</tr>
<tr>
<td>ICT officers</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Statisticians</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Education researchers</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Total</td>
<td>13</td>
<td>41</td>
</tr>
</tbody>
</table>

### IV. Results and Discussion

#### A. Results from the Developed Tool

The developed web-based tool allowed users to explore student dropout data in various levels and categories through different charts and maps. The first page of the developed tool displays visualizations for student dropout statistics in maps and donut charts. From this page, users can navigate through the tool using the tabs provided at the top of the page and selecting the category in which they want to see the visualizations. The tabs lead users into four main menus which
are Home, Secondary Schools, Primary Schools and Upload data. Figures 4-10 show parts of the developed tool.

Fig. 4. Maps showing dropouts in primary and secondary schools. Users can move the slider to view dropouts for different years and hover over a specific region to see the exact number. The donut charts below the maps present the dropout reasons.

Fig. 5. Grouped bar chart showing secondary school dropouts in each region by gender. In this case, the user has selected the secondary schools tab, then Male/Female dropouts and year 2016 from the dropdown list. The name and number of dropouts are displayed when a user hovers over a specific bar.

Fig. 6. A bubble chart showing primary school dropouts in each district of a selected region in relation to enrollment. The chart appears after a user selects a region and year from the dropdown lists.

B. Focus Group Discussion Results

During the focus group discussion, all participants generally reported that the data visualization tool was useful for getting a quick impression of the student dropout data. The majority liked the way data were presented in different kinds of charts and in various levels for regions and districts, though the majority suggested that more features should be added to the tool, especially going deeper to school level and including other education data apart from dropouts. To gather more specific information on the developed tool, the discussion was
led by a number of questions. The participants were asked whether they have used a similar tool before, which features they liked or disliked the most, when and how they would use the tool, and what they thought was the most important thing about the developed tool. A few participants stated that they have used a similar tool while the majority replied that they have not. Some commented further that even though they have not used a tool like this before, they considered it very helpful in making student dropout data understandable. On the feature preference, the majority liked the way data were presented in bar, pie and bubble charts. One participant commented that “The use of bubble charts helps to quickly compare the dropout rates among districts”. On the contrary another participant commented: “Why bubble charts? They are difficult to understand and they need someone to be familiar with them, unlike the other charts such as pie charts with numbers and percentages”. A few suggested the use of “bubbles over maps” instead of just bubble charts. This brings to light that, though data visualization has become inescapably important, some of the current visualization techniques are still unknown and that people prefer viewing data in visualizations that are familiar with. On using the tool, one participant commented: “The tool is very promising in visualization and can further be extended to present data about other situations apart from dropouts and be included in every system”. Another noted that, “It is very useful especially when researching issues related to education”. A further comment was that “It can be used when dealing with challenges in education” and “to intervene with dropouts and find a solution”. For the improvement of the tool for reporting and decision making, the participants suggested that the tool should incorporate other data, and present them down to school level. One participant noted that: “The tool needs to incorporate feedback to the actions taken, based on the data perceived in the presented visualizations”. The participants were also asked what was the most important thing regarding the visualization tool. Apart from the majority that said that the tool was simple and useful, one specifically said that: “The most important issue that I have come up with is that, it is easy for one to get information or data easily if tools like this are made or discovered”.

C. Questionnaire Results

The questionnaires for this study were filled on a 5-point scale from strongly agree to strongly disagree. The responses were centered on the ease of use, the usefulness of the tool at respondents’ line of work and at other departments, whether they will recommend it to their colleagues and if they think features should be added or removed. R statistical software was used to analyze the collected data form the questionnaires. On the ease of use, the majority (83%) agreed that the web based tool was easy to use and the rest (17%) strongly agreed that the tool was easy to use. Figure 11 shows these responses, describing also the position (title) of the respondents in a stacked bar chart. For the usefulness of the tool, 22% of the respondents strongly agreed that the tool was useful at their line of work, 56% agreed to it, while the rest 22% were neutral as shown in Figure 8. This discloses that even though interactive visualizations have been less used in student dropout issues in the context of non-online studying environment, they can yet be applied in this area to inform not only the decision makers but also the society in general. Moreover, 49% of the respondents strongly agreed that the tool would be useful to other domains and departments while 32% agreed and 19% were neutral as shown in Figure 13. The responses show that, visualizations play an important role for informing all domains of operation.
In checking if the tool would be recommendable, 7% of respondents strongly agreed that they would recommend it to their colleagues, 88% agreed that it was recommendable and the remaining 5% were neutral as shown in Figure 14.

![Figure 14. Responses on recommending the tool.](image)

Regarding feature addition, Figure 15 shows that the majority (59%) strongly agreed that more features should be added to the tool, the 39% agreed, while 2% of the participants were neutral.

![Figure 15. Responses on the need to add more features.](image)

D. Further Discussion of the Results

The results from focus group discussion and questionnaires confirm the usefulness of visualization for data presentation and informing decisions not only in school dropout issues but also in other areas of administration. However, the challenges encountered in data preprocessing due to the disparity of data enlightens the need for continuous and comprehensive data collection practices for the purposes of enabling good visualizations which will accurately inform the respective domain.

Moreover, the results apart from emphasizing on the importance of data visualization, also show that different users have different knowledge, understanding, and preferences on various kinds of visualization techniques, and that, in general, the visualization literacy is still low. Therefore, in order to exploit the benefits of data visualization, not only the data are needed, but also awareness of the users on the visualization techniques.

V. CONCLUSION AND FUTURE WORK

The developed visualization tool was generally agreed to be a simple, user-friendly, and recommendable way of presenting the large amounts of student dropout data. It was also agreed that it enhanced quick understanding and thus is useful for decision making and reporting. Nevertheless, the tool focuses on school dropout data which are only a very small subset of all educational data and presents them only up to council/district level while it does only annual reporting. So, further improvement is required in order to include deeper and wider levels of data and users. Moreover, information about other persistent global challenges in the education sector such as poor learning outcomes, equity issues, infrastructure and learning materials, quality of teaching, etc. could be included in the tool for making regional and global comparisons. This study sought to develop a web-based data visualization tool for student dropouts in Tanzania, an area which has been less explored and despite the limitations, it stands as a promising tool for informing all areas in education domain.

ACKNOWLEDGMENT

The authors would like to express their gratitude to the African Development Bank (ADB) for funding this study and the Nelson Mandela African Institution of Science and Technology (NM-AIST) in Arusha - Tanzania. The authors would also like to thank the heads of schools and the PO-RALG staff from the departments of Education, Statistics and IT.

REFERENCES
