

What Do I Do With All This Data? How To Use The FUN! Tool To Automatically Clean, Analyze, and Visualize Your Digital Data

Kevin Close, Utah State University
Phil Janisiewicz, Agile Dynamics
Sarah Brasiel, Utah State University
Taylor Martin, Utah State University, National Science Foundation

Abstract: Digital learning environments, particularly digital games, are becoming more prevalent, and datasets from these environments are becoming larger, but many researchers are still parsing data using spreadsheet software and graduate student labor. The academic community, like the business community, needs tools to automate repetitive processes and to manage the increasing amount of data from these digital learning environments. In this workshop, attendees will learn to use a new tool called the Functional Understanding Navigator! tool, or FUN! tool, to adapt, analyze, and present their digital data.

Introduction

Using data science methodologies to analyze digital learning environments has the potential to impact businesses and schools in big ways (e.g. personalized learning, learning pathway recognition, stealth assessment). The amount of time that workers and students spend in digital learning environments is increasing, resulting in larger, more complex digital datasets. The datasets are exciting, with the potential to reveal previously indistinguishable learning patterns and predict learner outcomes, but also frustrating for researchers with small teams and limited resources. Some problems include:

1. Coordinating research teams consisting of personnel from several departments.
2. Efficiently managing large amounts of data.
3. Securely managing large amounts of data.
4. Reproducing results after several iterations of data cleaning and analyses.

The Functional Understanding Navigator!, or FUN! tool, addresses this pressing need for tools that automate processes to help researchers efficiently clean, analyze, and present data (see Figure 1). Furthermore, the FUN! tool defines a data pipeline to keep data more organized. The tool also logs the history of the data analysis process so the research team can easily trace their steps and reproduce their results.

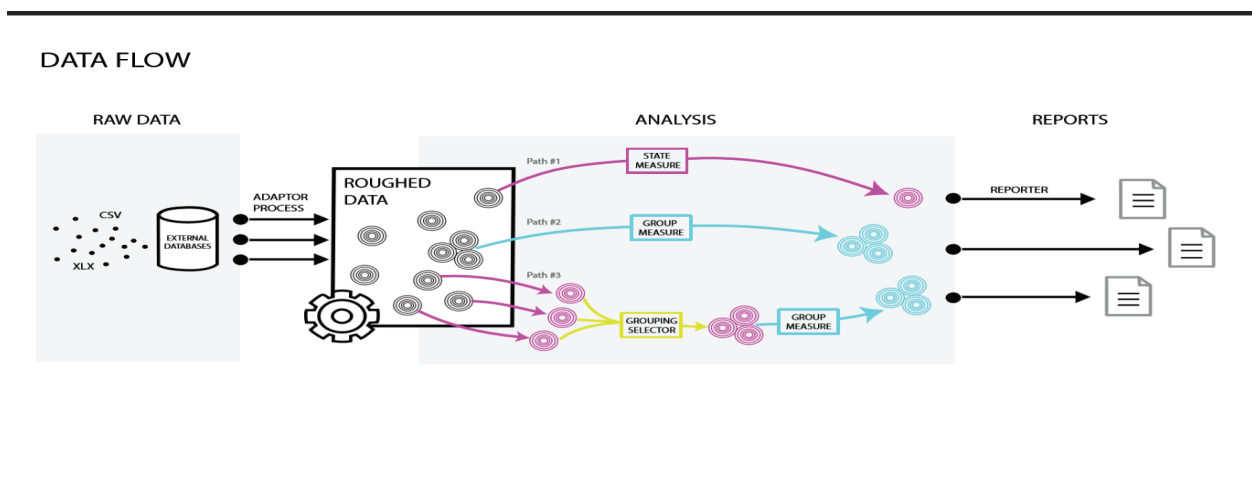


Figure 1: The FUN! tool has three distinct steps: adapt, analyze, and report data.

Introduction (10 minutes)

The workshop organizers began with an introductory discussion about the current trends and issues in educational data mining and learning analytics. We provided an overview of the type of research questions and data that are most relevant to our methodologies and the tools that are currently available for researchers, which will help contextualize the FUN! tool. We also discuss the importance of building a relationship with developers during the data collection process. Before beginning the hands on exploration, we summarized three key points as to the value of the FUN! Tool: (a) organizes your workflow process from start to finish, (b) records log data of all of your analyses, and (c) provides a platform to share your analyses with others.

Hands on Exploration of the FUN! Tool (40 minutes)

Next the workshop organizers provided a 40-minute hands-on workshop for participants to explore the FUN! Tool using laptops at six stations loaded with the FUN! Tool and practice data. A User Guide was provided to all participants with directions to follow along with during the exploration and to use with their own data. During this session, attendees worked with a sample dataset from the *Scratch* programming language. Specifically, attendees learned to: (a) modify the adaptor of the FUN! tool (so it can read in their own data), (b) modify the analysis functions (so the tool can select various data pieces and measure various aspects of those data pieces), and (c) modify the reporters (so researchers can read the results).

Discussion (10 minutes)

The workshop concluded with a discussion about how to use the FUN! tool with various datasets. We asked the audience to discuss the value they recognized with use of the FUN! tool. Several people seemed to see the value in using the FUN! Tool to automate processes done routinely and as a way for data scientists to support the work of researchers. The participants were not familiar with *Scratch*, so they appreciated the discussion of the application of the FUN! tool to another environment that uses log data of moves and board states from a fraction game. This data seemed similar to the data many of the participants worked with and allowed them to see the potential of the FUN! tool processes for use with their own data. We had a discussion of recommended next steps. Some felt the development of a dashboard was important to truly visualize the data. Others felt that more time could be spent on measure development and that the dashboard could be left to individual users and their needs for what should be visualized.

Conclusion

Though automated analysis tools, like the FUN! tool, are already being used in the business sector (see *Spark* and *KNIME*), the FUN! tool is part of the first generation of tools to anticipate and address the challenges that will come with the incoming data deluge in education (for other examples see Stenerson, Salmon, Berland, & Squire, 2014 and Werner, McDowell, & Denner, 2013). We envision tools like the FUN! tool, to eventually be used in training environments and classrooms so that instructional designers and teachers can automatically analyze learning environments and make better formative decisions.

References

- Stenerson, M. E., Salmon, A., Berland, M., & Squire, K. (2014, October). ADAGE: An open API for data collection in educational games. In *Proceedings of the first ACM SIGCHI annual symposium on Computer-human interaction in play* (pp. 437-438). ACM.
- Werner, L., McDowell, C., & Denner, J. (2013). A first step in learning analytics: Pre-processing low-level Alice logging data of middle school students. *Journal of Educational Data Mining*, 5(2), 11-37.

Acknowledgments

This work is supported by National Science Foundation Grant IIS-1319938. Any opinions, findings, conclusions or recommendations expressed in this material and presentation are those of the authors and do not necessarily reflect positions or policies of these agencies. This material is based upon work supported by and while the author, Taylor Martin, was employed by the National Science Foundation. Any opinion, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.