CSC591/791: Assignment 2

August 29, 2017

Due 9/5/2017 8:30am.

In this assignment you will gain some basic experience with data munging and data visualization. In this assignment you will select one of the Andes datasets from the PSLC DataShop (`https://pslcdatashop.web.cmu.edu/Project?id=27`). Andes is an Intelligent Tutoring System for Physics that provides students with automated guidance as they solve physics problems. The function of Andes is described in the following paper:

> Kurt VanLehn et al. (2005). "The Andes Physics Tutoring System: Lessons Learned". In: *International Journal of Artificial Intelligence in Education (IJAIED)* 15.3, pp. 147–204. URL: `http://people.engr.ncsu.edu/cflynch/Papers/AndesLessonsLearnedForWeb.pdf`. **(cited 531 times)**

Tasks

For this assignment you will:

1. Select an Andes dataset from the link.
2. Segment the individual data into per-student records covering all problem solving.
3. Calculate the total time spent for each student.
4. Calculate the number of actions performed per minute.
5. Produce a visualization of each value.

All of the data processing and generation of the visualization should be done in code. You should submit the dataset, your code, and a written technical report that describes the what you did, how to use your code, and show the visualizations that you developed.

Due to the complexity of the assignment you are allowed to work in a pair. Only one individual should submit the code and the report. The final report must clearly list both authors.

Logfile Format

The Andes data consists of a series of concatenated log files. Lines in the log files are header lines, comment lines (preceded by a #) or are log lines which include both a timestamp, and individual action information.

For the purposes of this assignment you can focus on the DDE and DDE-POST lines which represent individual actions. A student record is loaded with a line of the form:

```
0:01 DDE (read-student-info "1D425" 1)
```

Opening a new problem occurs with a line of the form:

```
0:01 DDE (read-problem-info "FOR1B" 0 0)
```

And problems are closed with the command:

```
11:59 DDE (close-problem "GAUSS3")
```

You may safely ignore all other lines apart from DDE and DDE-POST. You are encouraged to read through the files to identify the structure.

Submission

You should submit three separate files:

- The dataset used (`YourName_dataset.zip`).
- Your code for processing the dataset (`YourName_code.zip`).
- Your written report (`YourName_report.pdf`).