Massive Open Online Course (MOOC) Analytics and Visualization

Howard Baek, Huzefa Rangwala, Jessica Lin, Jill K. Nelson

howardba@uw.edu
University of Washington

rangwala@cs.gmu.edu
George Mason University

jessica@gmu.edu
George Mason University

jnelson@gmu.edu
George Mason University

This work is funded by NSF REU Site on Educational Data Mining at George Mason University; Grant No. 1757064

Background:

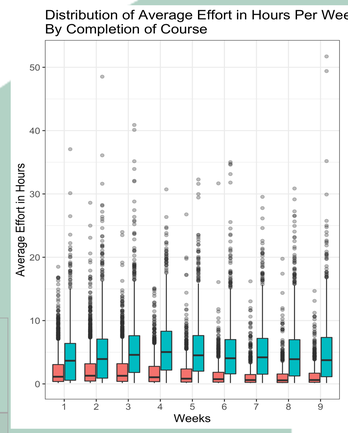
- MOOCs offer low-cost and convenient education to students around the world.
- Huge problem in MOOCs is the low-completion rates
- Widely cited dropout rate is 90%

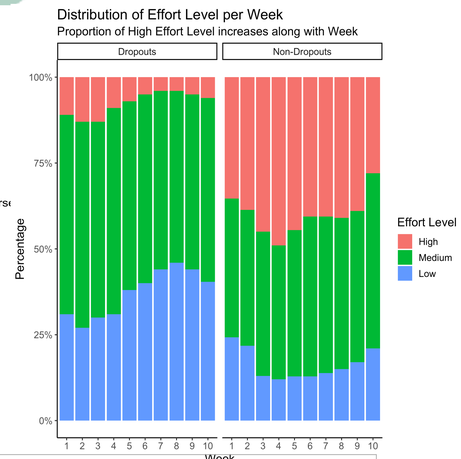
Objectives:

- Understand datasets on Stanford edX's Statistics in Medicine course.
- Employ clustering to identify factors leading to dropouts
- Provide a interactive web application for instructors and course administrators

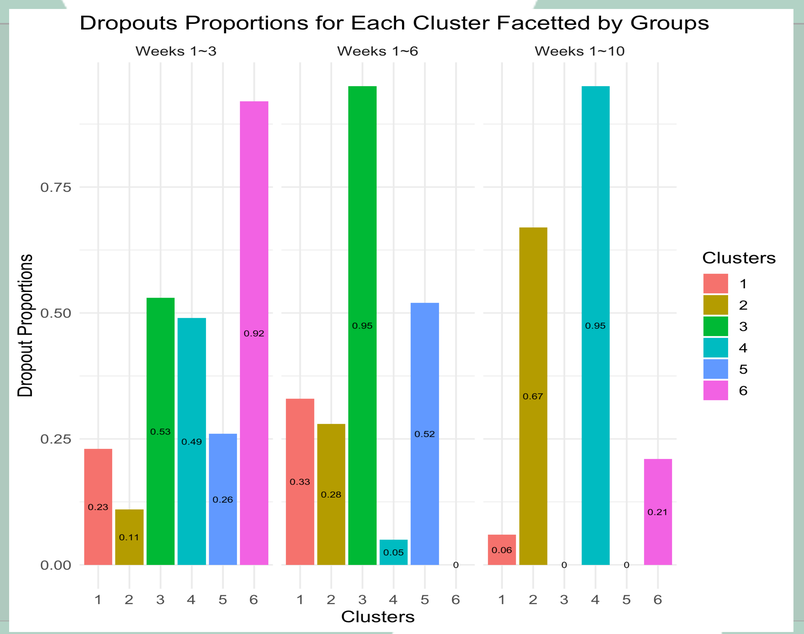
Exploratory Data Analysis:

- Number of students who are in all four datasets: 7659
- Number of students who dropped out of the course: 4996
- Dropout Rate: 65%





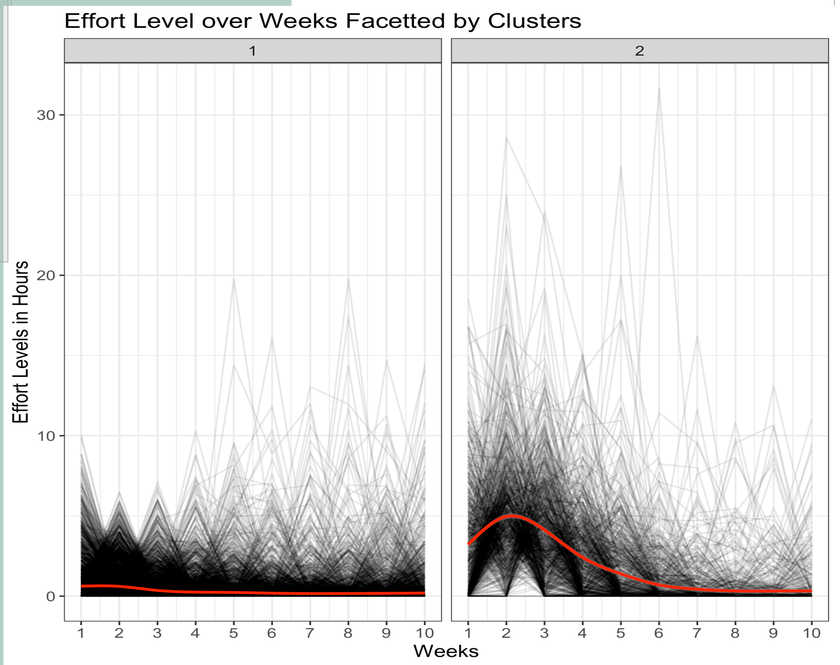
Clustering (K=6)



	Accuracy	Precision	Recall	F1 Score
Full	0.72	0.74	0.89	0.81
Halves	0.79	0.77	0.97	0.86
Quarterly	0.80	0.78	0.98	0.87

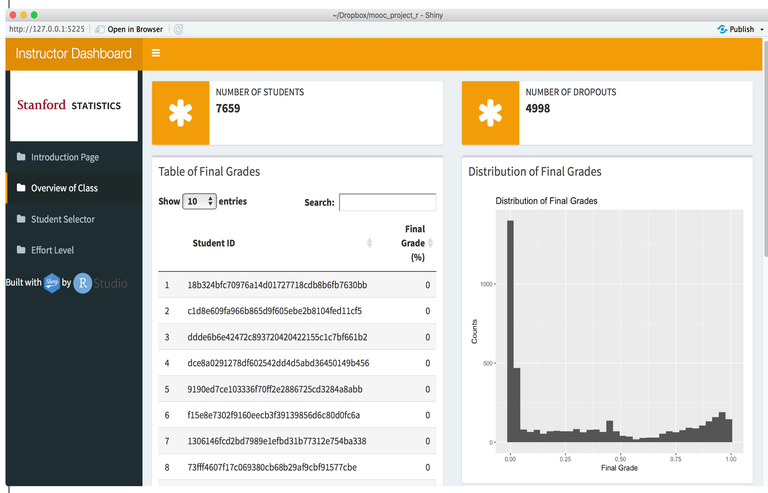
Clustering on Dropout Students

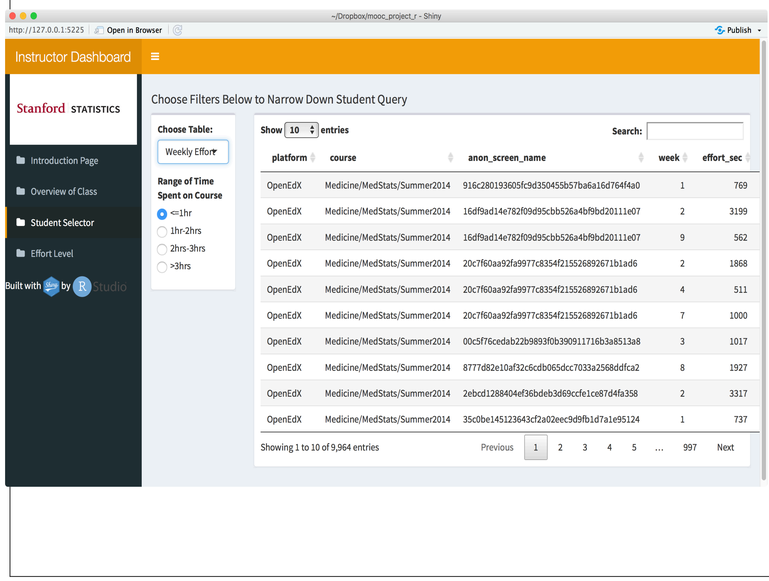
- Defined dropouts as students who watched fewer than 50% of videos.
- Among dropouts, observed two clusters of students with distinct trend lines of effort levels



MOOC Instructor Dashboard

- Intended for teachers and administrators to visualize and understand the data.
- Built using shiny and shinydashboard R packages.





Code is available at https://github.com/howardbaek/mooc-project-github