MOOC Visual Analytics: Empowering Students, Teachers, Researchers, and Platform Developers of Massively Open Online Courses

Scott R. Emmons
Robert P. Light
Katy Börner

Abstract

Massively open online courses (MOOCs) offer instructors the opportunity to reach students in orders of magnitude greater than they could in traditional classroom settings, while offering students access to free or inexpensive courses taught by world-class educators. However, MOOCs provide major challenges to teachers (keeping track of thousands of students and supporting their learning progress), students (keeping track of course materials and effectively interacting with teachers and fellow students), researchers (understanding how students interact with materials and each other), and MOOC platform developers (supporting effective course design and delivery in a scalable way).

Along with these challenges, the sheer volume of data available from MOOCs provides unprecedented opportunities to study how learning takes place in online courses. This paper explores the use of data analysis and visualization as a means to empower teachers, students, researchers, and platform developers by making large volumes of data easy to understand. First, we introduce the insight needs of these four user groups. Second, we review existing MOOC visual analytics studies. Third, we present a framework for MOOC data types and data analyses to support different types of insight needs. Fourth, we present exemplary data visualizations that make data accessible and empower teachers, students, developers, and researchers with novel insights. The outlook discusses future MOOC opportunities and challenges.

Data

Demographic Data
General student demographics, including age, gender, language, education level, and location. Demographic data is commonly acquired during the registration process, and additional demographic data can be acquired via feedback surveys.

Performance Data
Student performance based on graded assessments. This is generally collected from homework, quizzes, and examinations, but it also includes results from pre-course surveys designed to examine student knowledge before they take the course.

Activity Data
How students are using class resources, such as the time and date of watching videos, reading material, turning in homework, taking quizzes, or using the discussion forum. Most platforms break down usage by content and media type (i.e. page views, assignment views, textbook views, video views). Path through content via inbound and outbound links is important for understanding learning trajectories.

Feedback Data
Student input and feedback. Feedback data allows course providers to learn more about student learning goals and motivation, intended use of course, and content hoped to learn. It data also contains information about what students liked or disliked in terms of course content, structure, grading, and teacher interaction.

Data Type	Data Field	edX	Canvas	Coursera	GCB
Demographics	Email	D	DE	DE	DE
	Gender	D	X	DE	X
	Age / Birth Year	D	X	DE	X
	Location	D	X	D	GA
	Level Education	D	X	D*	X
Performance	Class Grades	D	D	D	D
	Quiz / Question Breakdown	D	D	D	D
	Student Breakdown	D	D	DE	DE
Activity	Test / Assignment Completion	D	D	D	D**
	Content Usage Breakdown	DE	D	D	GA
	Path Through Content	DE	DE	DE	GA
	Time-stamped Activity	DE	D	DE	GA
	"Active" Student Count	D	D	D	X
	Student Breakdown	DE	D	DE	X
Feedback	Supports Surveys	✓	✓	✓	✓

D = Dashboard DE = Data Export GA = Google Analytics

*If students choose to take optional demographic survey
**If toggled to record

Acknowledgements

We would like to thank Samuel T. Mills for re-designing the figures in this paper and all 2013 and 2014 IVMOOC students for their feedback and comments, enthusiasm, and support. All R code and all Sci2 Tool workflows are available and documented at cns.iu.edu/2015-MOOCVis.

Analysis Types vs. User Needs

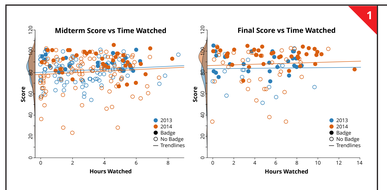
A. Student
Students taking MOOCs need to be extremely organized and disciplined. MOOCs have no weekly in-class teacher encounters. MOOCs also have much less peer-pressure. Courses might have vastly different schedules, activities such as labs and capstone projects, deadlines, and grading rubrics. Effectively using one or more MOOC platforms can itself be a major learning exercise. Plus, many students are not used to collaborating with students from different disciplinary backgrounds and cultures that speak different native languages and live in different time zones.

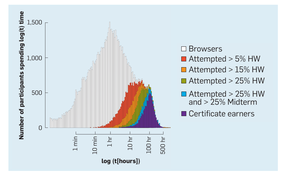
B. Teacher
Teachers of MOOCs need effective means to keep track of and guide the activities, progress, and any problems encountered by thousands of students. They need to understand the effectiveness of materials, exercises, and exams with respect to learning goals in order to continuously improve course schedules, activities, and grading rubrics.

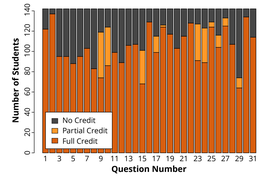
C. Researcher
Researchers that study human learning are keen to understand what teaching and learning methods work well in a MOOC environment and now have massive amounts of detailed data with which to work. As all student interactions—with learning materials, teachers, and other students—are recorded in a MOOC, human learning can be studied at an extreme level of detail. Many MOOC teachers double as learning researchers as they are interested to make their own MOOC course work for different types of students.

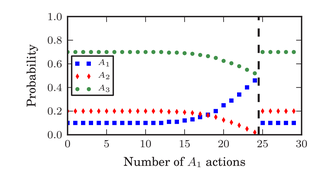
D. Platform Developer
Platform developers need to design systems that support effective course design, efficient teaching, and secure but scalable course delivery.
They need to support times of high traffic and resource consumption and schedule maintenances during low activity times.

1. Statistics
Line graphs, correlation graphs, and box-and-whisker plots are all examples of how statistical data can be rendered visually.

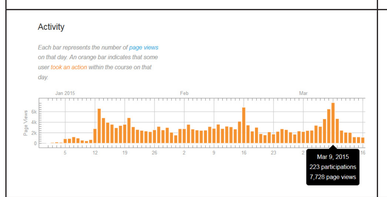


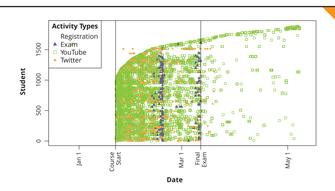


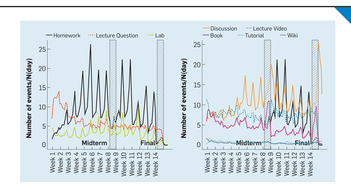


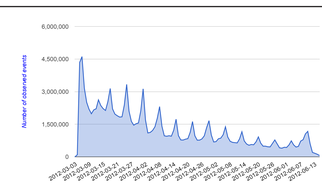


2. Temporal
Temporal analyses and visualizations tell when students are active over the span of a course. Data might be examined at different levels of aggregation: by minute, hour, day, week, or semester, by course modules, or before and after a midterm or final.

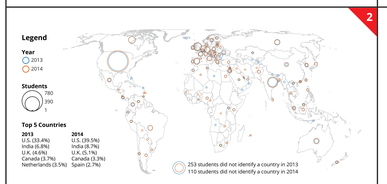


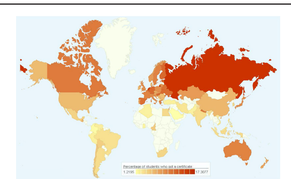


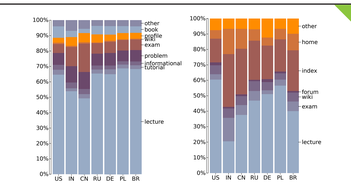


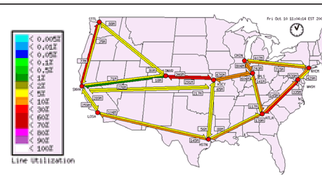


3. Geospatial
Geospatial data might be examined at different levels of aggregation: by address, city, country, or IP address.

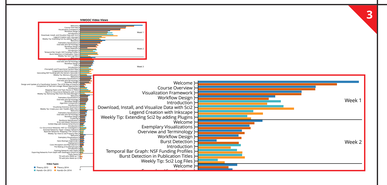


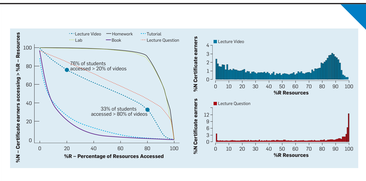


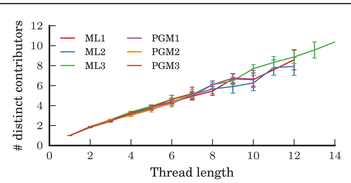


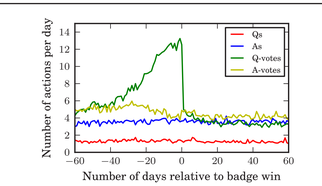


4. Topical
Topical analysis provides an answer to the question of "what" is going on in a course.









5. Network
Student cohorts might be created based on prior expertise, geospatial region or time zone, access patterns, project teams, or grades.