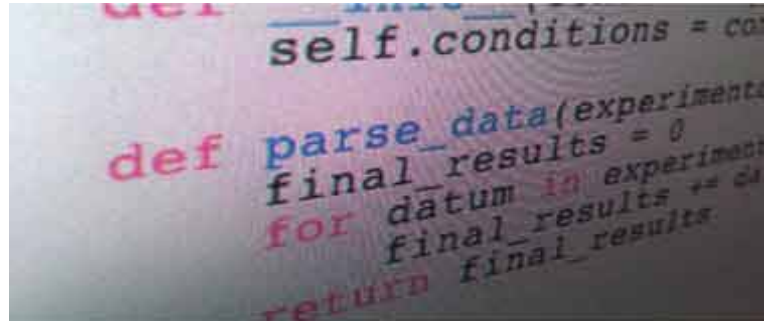


# 6.00x Introduction to Computer Science and Programming MITx on edX – 2012 Fall



“6.00x is an introduction to computer science as a tool to solve real-world analytical problems.” – edx.org

**Introduction.** 6.00x is an introduction to using computation to solve real problems. The course is aimed at students with little or no prior programming experience who have a desire (or need) to understand computational approaches to problem solving. Some of the people taking the course will use it as a stepping-stone to more advanced computer science courses, but for many, it will be their first and last computer science course.

## Inside

1. Introduction
2. Course structure
3. Demographics
4. Participants
5. Activity
6. Persistence
7. Special Features

6.00x was offered as an MITx on edX course in the Fall of 2012, taught by Professors Eric Grimson and John Guttag, and Dr. Chris Terman.

Many participants in MITx on edX courses come with different intentions and levels of interest and many resemble shoppers and auditors in their course activity. This report identifies four subpopulations of participants (numbers below): those who registered, viewed some course content, explored the course, and were certified<sup>1</sup>.

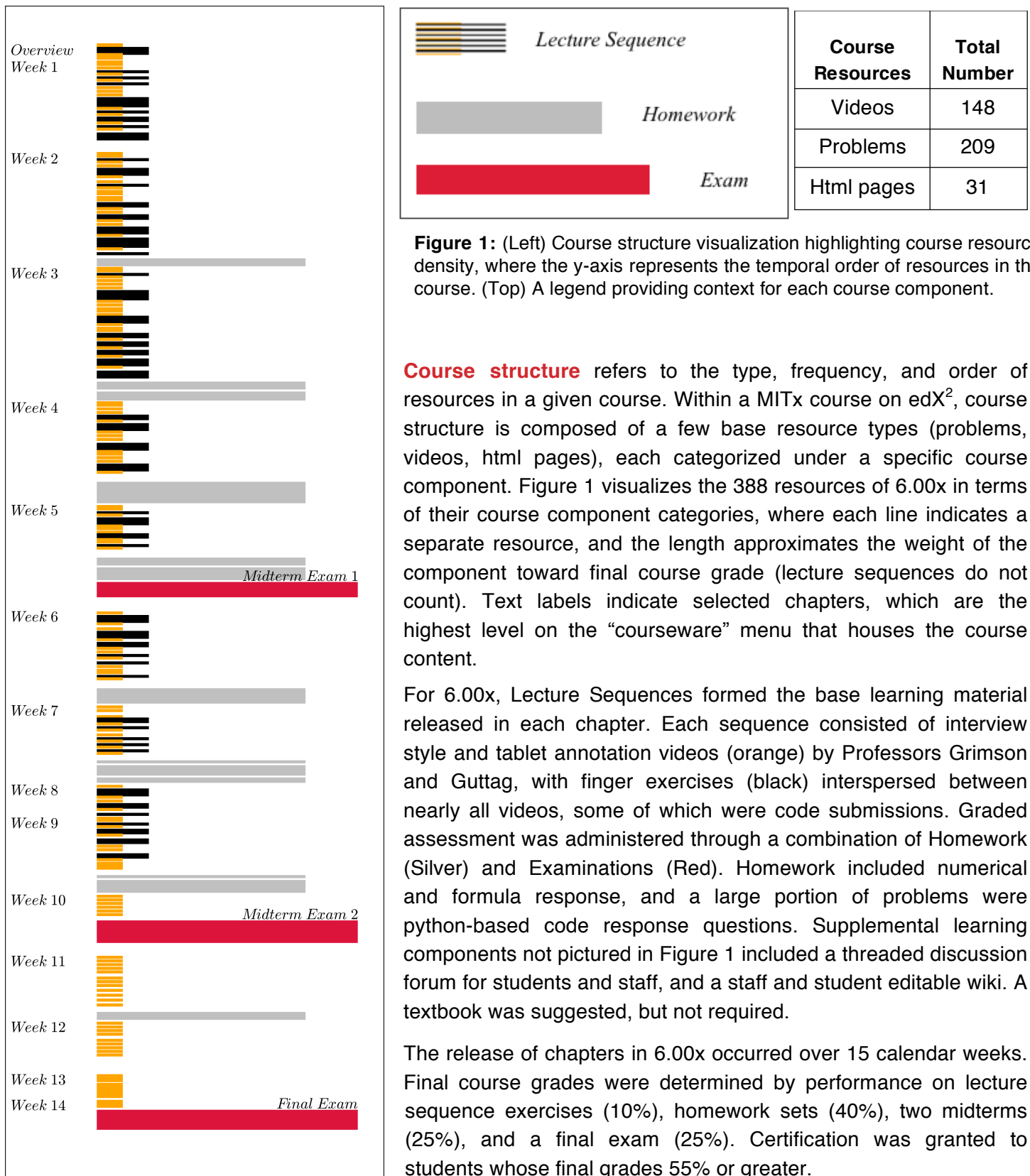
Important Dates	
July 24, 2012	Registration Open
September 26, 2012	Course Launch
January 15, 2013	Course Close
September 8, 2013	Date of Report Data Collection

Participants	
Registered	84,511
Viewed	55,339
Explored	8,941
Certified	5,732

<sup>1</sup> See page 4 for definitions

This report was produced jointly by the MIT Office of Digital Learning, and the HarvardX Research Committee.

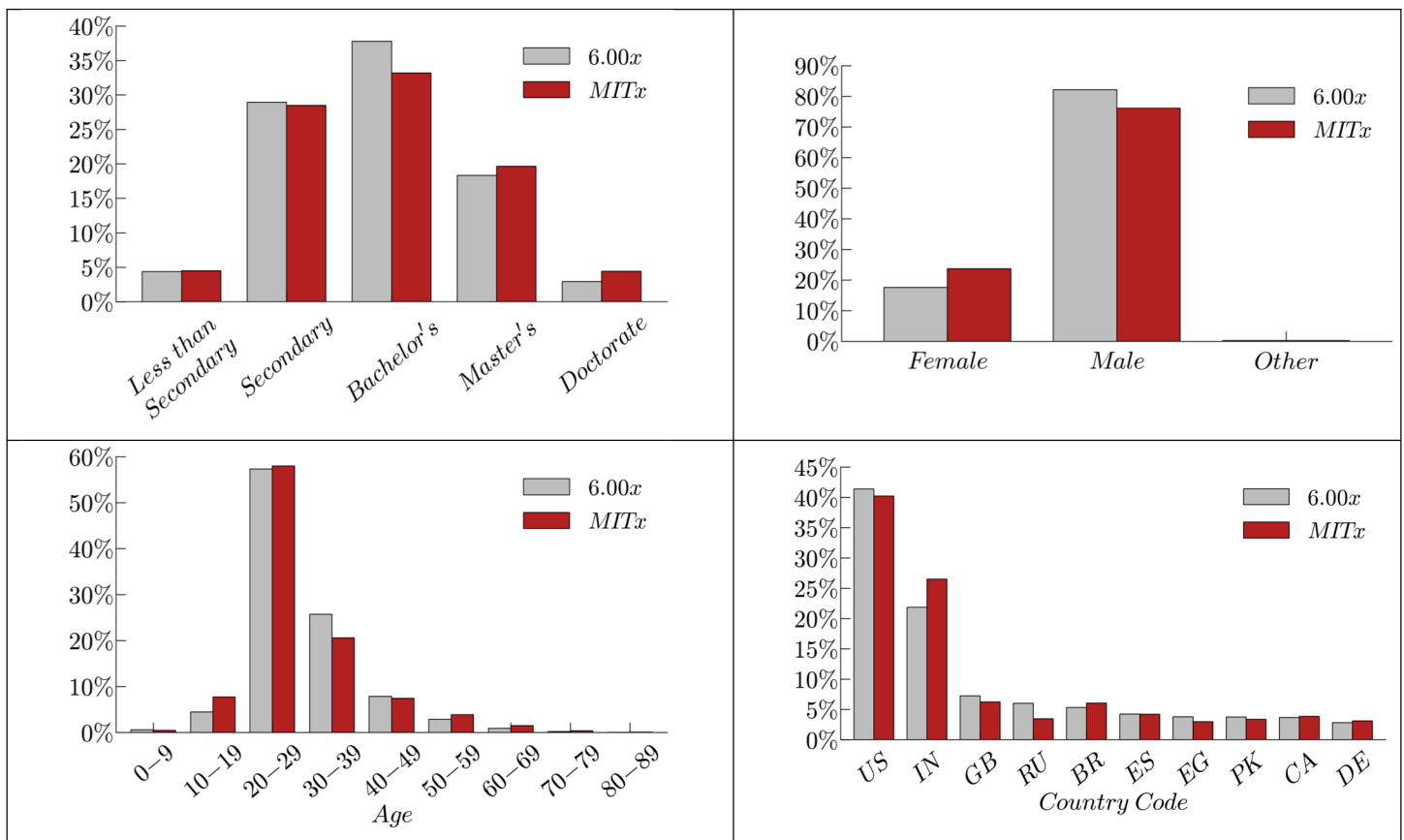
Seaton, D. T., Reich, J., Nesterko, S, Mullaney, T., Waldo, J., Ho, A., & Chuang, I. (2014). 6.00x Introduction to Computer Science and Programming - Fall 2012 MITx Course Report (MITx Working Paper #3)



<sup>2</sup> Often colloquially referred to as a Massive Open Online Course (MOOC)

To date, MITx courses on edX have had similar schedules of regularly released resources and periodic due dates. Like 6.00x, courses have used quantitative assessment supported by lecture sequences and textual material, as well as a few special learning components (see end of this report). In the figures that follow, we compare 6.00x data with those from other MITx courses on edX, from the Fall of 2012 to the Summer of 2013.

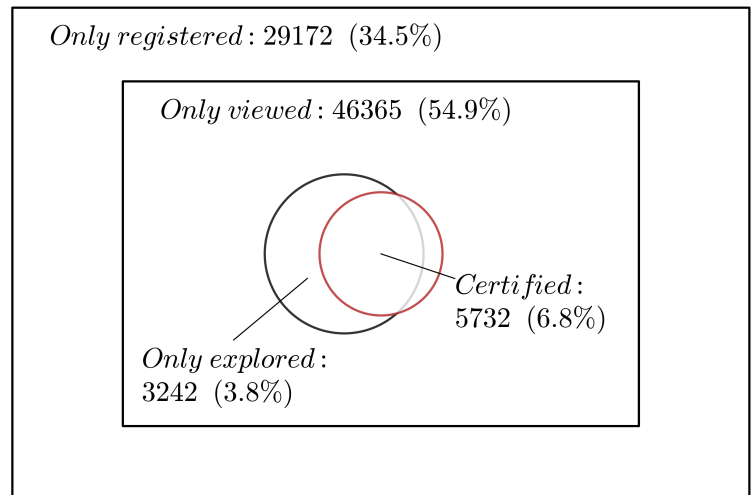
**Demographics.** For 6.00x, we have records of 84,511 students who registered for the course by the cutoff date of this report, September 8, 2013. According to self-reported demographics, registrants were 82.1% men, and as a group, highly educated: 40.9% have earned bachelor’s degrees, 19.8% have masters or professional degrees, and 3.2% of registrants had earned a doctorate (for certified students, percentages are 34.7%, 28.8%, and 4.8%, respectively). Distributions of level of education and gender are plotted in Figure 2 below, along with the average distributions for all MITx courses<sup>3</sup>. Note, roughly 5% of 6.00x registrants did not self-report one or more of the demographic categories.



**Figure 2:** Distributions of demographic variables collected at registration for all registrants reporting data, with MITx-wide comparisons for reference: (top left) education attainment, (top-right) gender, (bottom left) age, and (bottom right) geolocation via IP look-up.

<sup>3</sup> See also Ho, A. D., Reich, J., Nesterko, S., Seaton, D. T., Mullaney, T., Waldo, J., & Chuang, I. (2014). *HarvardX and MITx: The first year of open online courses*. (HarvardX and MITx Working Paper No. 1).

As with other MITx courses on edX, 6.00x students came from all over the world. Based on geolocation of student IP addresses, 25.1% of registrants came from the United States, but many students also came from over 150 other countries around the world. Consistent with other MITx courses on edX, the median age of registrants and certified students were higher than those of typical MIT undergraduates; namely, 27 or 28 years, respectively. Distributions of geolocated country and age are plotted in Fig. 2 for 6.00x, along side the average distributions for all MITx courses.



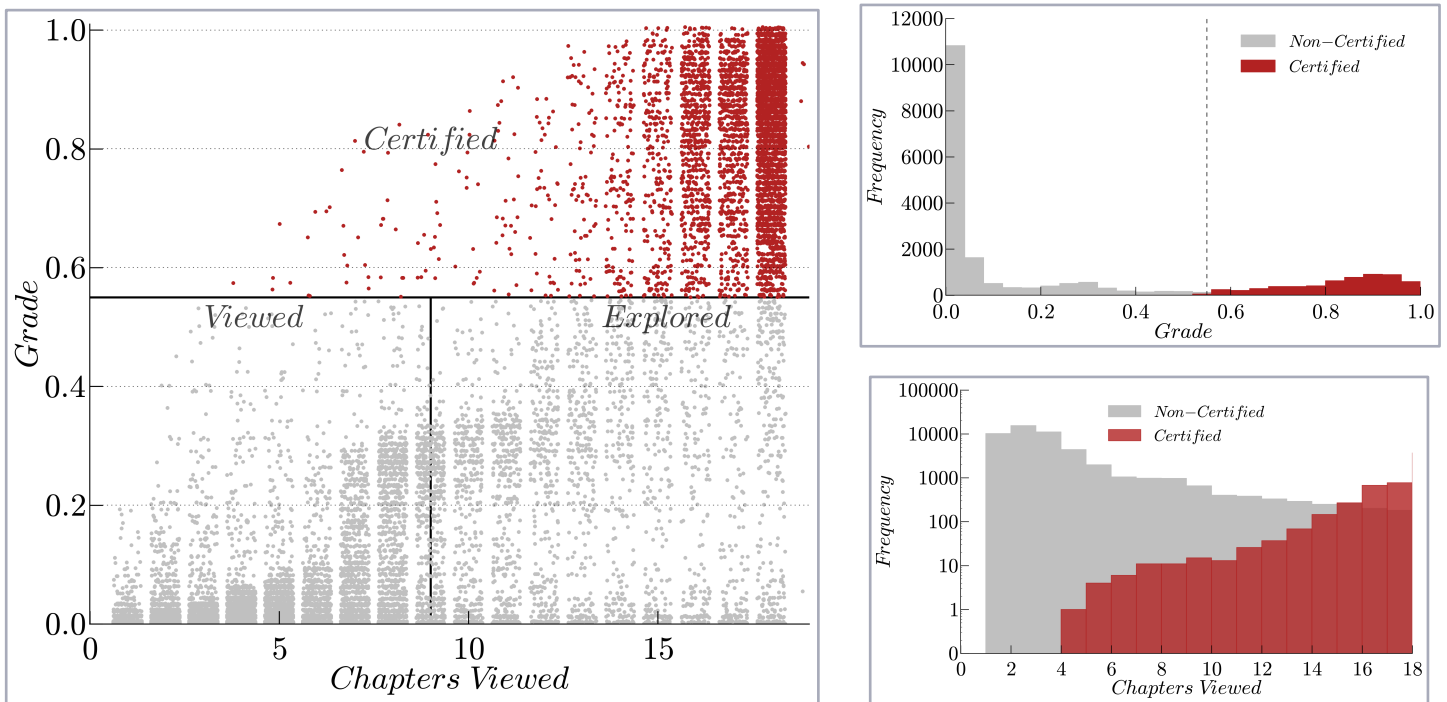
**Figure 3:** Participants separated into four mutually exclusive and exhaustive categories (not to scale).

**Participants in 6.00x.** Who are the students taking this course, and how much did they participate? The degree and kind of participation in the course varied considerably among those who signed up for 6.00x. Furthermore, like many MOOCs, this course remained open to new registrants, so that enrollment continued to rise through the semester. This asynchronicity is a key feature of open online courses. One feature of low barriers to registration, for example, is a large number of students who registered but never actually viewed any course material.

To illustrate the considerable variability in participants by their actions, we identify four subpopulations of interest within this course: those who (1) only registered, and never accessed the courseware, (2) only viewed, i.e. non-certified registrants who accessed at least one chapter but less than half of chapters, (3) only explored, i.e. non-certified registrants who accessed half or more of the chapters, and those who were (4) certified, by earning a final grade at or higher than the cutoff.

Certificate earners are the most conventional subpopulation, defined by crossing a certain threshold of achievement on quantitative assessments (well defined for 6.00x). The “only explored” subpopulation seeks to represent students having interacted with a considerable amount of course material, while the “only viewed” subpopulation highlights participants having only sampled the course. Both the “only explored” and “only viewed” subpopulations contain participants with highly varying levels of activity (as highlighted in the following two sections on persistence and activity). The fourth sub-population of “only registered” contains users that registered, but never accessed the courseware. This is a substantial number in 6.00x and many of the MITx courses on edX.

For 6.00x, the numbers of students who only registered, only viewed, only explored, or were certified, are shown in Fig. 3. The categories are defined as disjoint sets by using the “only” terminology, but the figure highlights the nested nature of registration, viewing, exploring, and certification. Students who were certified

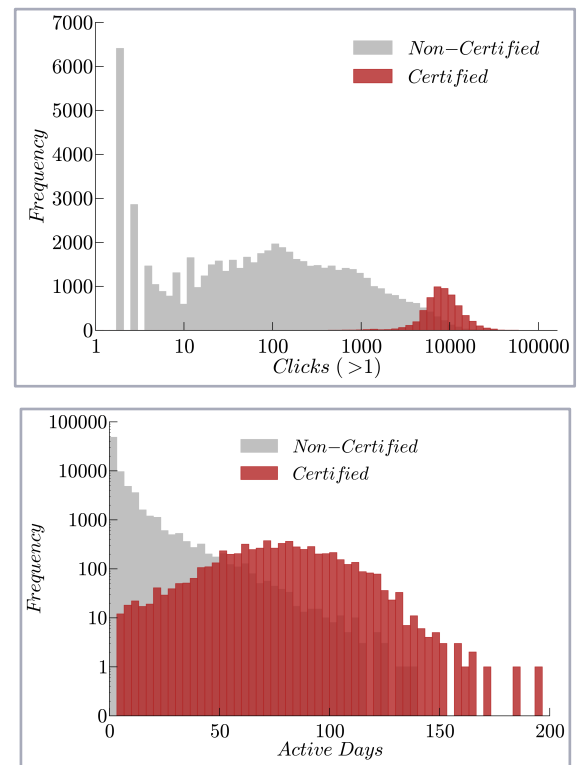


**Figure 4:** Scatter plot of grade versus chapters viewed (left), highlighting student sub-populations; certified students are red points and all points are jittered. Histograms of grades and number of chapters viewed (right) distinguished by student certification status.

certainly viewed the course, but not all viewers were certified. Similarly, “explored and certified” are both subsets of viewed.

**Activity in 6.00x.** The activity of students in MITx courses on edX is logged; activity level can provide perspective into which courseware components are of interest to students, and which specific activities students are undertaking in a course, e.g. to earn certification.

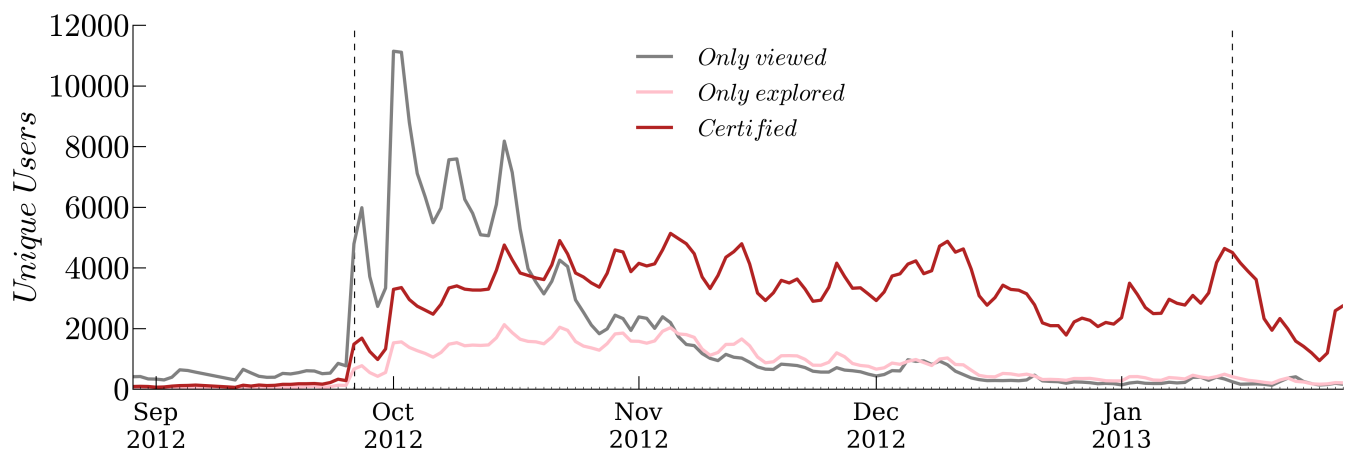
How diverse is the activity of students in the course? Consider the grade earned by students, which is determined in 6.00x by quantitative assessment with regular due dates and periodic exams. A scatter plot (Fig. 4) depicting earned grade versus the number of chapters visited, for each registered student, illustrates the distinctions between the sub-populations of students who only viewed, only explored, or became certified in the course. This plot also illuminates some possible registrant types: “completionists” (perfect grade, viewed all chapters), “optimizers” (occasional



**Figure 5:** Distribution of clicks (total course interactions) and active days, distinguished by certificate status.

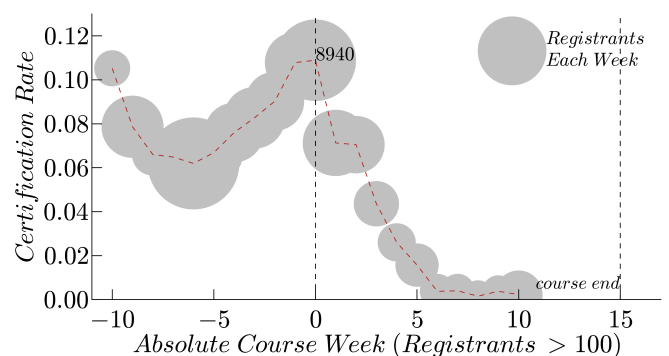
registrants that view minimal chapters but earn a certificate), and “listeners” (viewed all chapters, zero grade – submitted no work for credit).

For 6.00x, edX recorded 89,650,395 events in its tracking logs, relating to students interacting with the course contents, through the cutoff date of this report. These events included interactions indicating when students played videos, attempted problems, browsed through text pages, read or posted entries in the forum, and other activities. Fig. 5 gives a glimpse into these data, presenting a histogram of the total number of events (“clicks”) recorded for all registrants with at least one event, and highlighting the population of students who earned a certificate in comparison with all other students. The figure also presents a histogram of the number of active days, where “active days” represents any day where a course interaction was logged. The distribution of active days for certificate earners is broad, further reflecting the variety of approaches by which a participant can earn a certificate.



**Figure 6:** Daily number of unique participants accessing courseware or course-site (pre-launch) via our described participant categories. Vertical dashed lines indicate the start and end of the course.

**Persistence in 6.00x.** How does the activity of students change as the course progresses, particularly with students who persist to earn certification? Figure 6 highlights courseware and course-site (pre-launch) interactions for three of the disjoint participant types: only viewed, only explored, and certified. In 6.00x, there is a noticeable periodicity in each curve that is highly correlated with content release and due dates. Certified students show a relatively stable weekly periodicity, while the periodicity of the viewed and explored curves disappears after about one-third and two-thirds of the course, respectively. It is important to note that these curves do not account for registration date or for large times between user interactions. However, they do provide insight into the collective



**Figure 7:** Certification fraction versus registration week relative to course launch. Bubble size represents the number of unique users registering each week.



behavior of our defined populations. It is particularly interesting to formulate hypotheses about behavior early in the course, e.g., attempting to target struggling students that may benefit from intervention.

Do students who register early have a higher likelihood of earning a certificate than students who register late? Figure 7 depicts the certification fraction versus registration week relative to course launch (bubbles indicate total enrollment each week). Across MITx courses, certification rates do not depend greatly on registration dates prior to launch, although some courses have higher registration rates very early as well as just prior to launch. For 6.00x, between 6% and 12% of users registering prior to launch earn a certificate regardless of enrollment week. After launch, the rate drops dramatically. Because of the highly structured due dates in 6.00x, it is reasonable to attribute most of the post-launch drop in certification rates to registrants who have missed due dates and find it difficult or impossible to catch up.

### Special Features

Each MITx on edX course includes special features unique to that course. In particular, 6.00x offered assessment questions whose answers were student submitted computer programs (python code), with the following features:

- **Instant assessment:** Programs are graded and feedback given to students typically within a few seconds after submission.
- **Functional correctness evaluation:** Student code is checked for functional correctness, using unit tests designed by the instructors. Students are instructed to write code which has a certain behavior; instructors provide “golden master” code representing a correct answer, and the grading system runs both, with the unit tests. Test results are compared, and shown to the student. For example, one common problem asks students to write a program to compute a Fibonacci sequence, from an intermediate point. The unit tests start the program at random points, comparing the golden master’s output with the student’s code output. The detailed feedback provided to students help guide code revisions and clarification of thinking.
- **Style correctness evaluation:** The structure of the code is analyzed for certain properties specified as needed by the instructor. For example, the instructor may ask students to solve a given problem using recursion instead of iteration. The text and syntax tree of the student’s code is analyzed to check for such stylistic correctness.
- **Code style evaluation:** Good coding principles go beyond functionality, to include ideas such as not overloading variables, implementing docstrings, and employing local instead of global variables. Such stylistic properties are checked by a special piece of grading code, which presents text feedback to students, and allows instructors to deduct points for style errors.