MOOC En Images
6.002x

Franck Dernoncourt, Colin Taylor,
Kalyan Veeramachaneni, Una-May O'Reilly

ANY SCALE LEARNING FOR ALL GROUP
Computer Science and Artificial Intelligence Laboratory
Massachusetts Institute of Technology

Table of Contents

List of Figures

Preface

This report provides a view into different descriptive statistics extracted from the data recorded during 6.002x the first course offering by MITx. We have developed a generalizable analytics framework and this report demonstrates use of this framework.

July 26, 2013.

1 Introduction

We have developed a comprehensive data schema that captures the data from a MOOC. The data schema and the rationale behind the schema is presented in [1]. Building a data schema enables generation of multiple views of the data. Providing views into the data enables several stake holders to quantitatively measure either retrospectively or during the course as to what happened in the course. These views we call "MOOC En Images". We note that there are fundamentally four different axes in this data. They are: *the student*, *time*, *space*, *course*. These axes define how an viewer would like to dissect the data and look for a pattern. For example a query could be "*The total number of submissions per week from the students who got an A grade in the course*". In this query the user has cut the time axes on a weekly basis, selected a student cohort that got an A, defined an aggregate statistic - total number of submissions. The user however did not define any cut for the course axes. However, if the query is slightly modified as follows "*The total number of submissions per week, for homework 3, 5 and 7, from the students who got an A grade in the course*" defined specific sub components in the course.
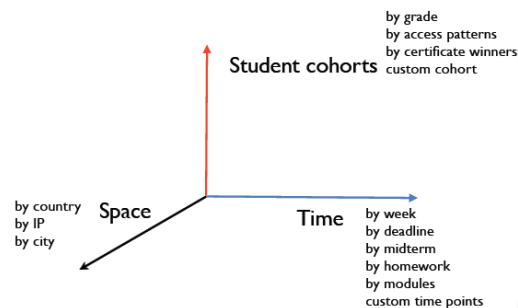

Fig. 1. The 3 main axes.

We have populated the schema with the data from 6.002x. In this report we present several views into the data and present how an user can interface with our system and create customized views from a courses data.

2 Grade Distribution
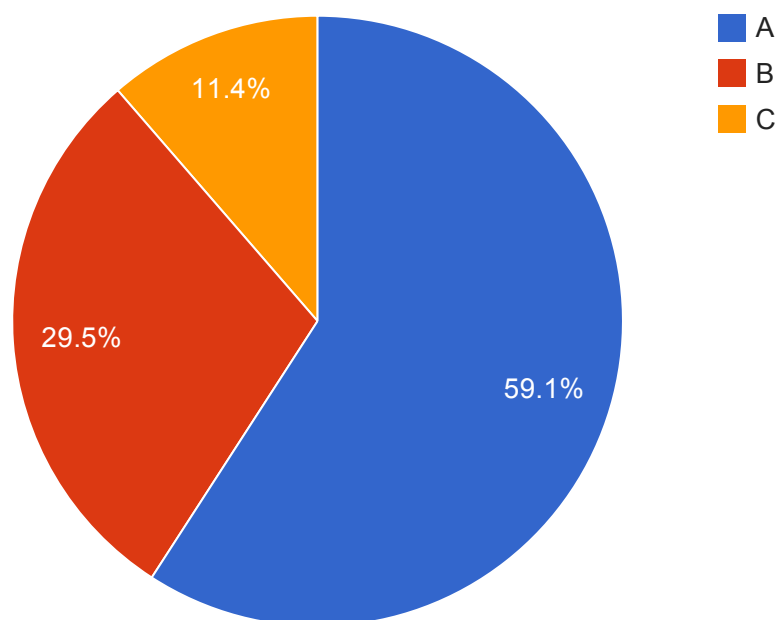
Final grade distribution

Fig. 2. Final grade distribution of students who received a certificate in 6.002x. Students who earned at least 60% received a C, at least 70% received a B, and who earned at least 87% received an A. Over 59% of students who received a certificate got an A in the class.
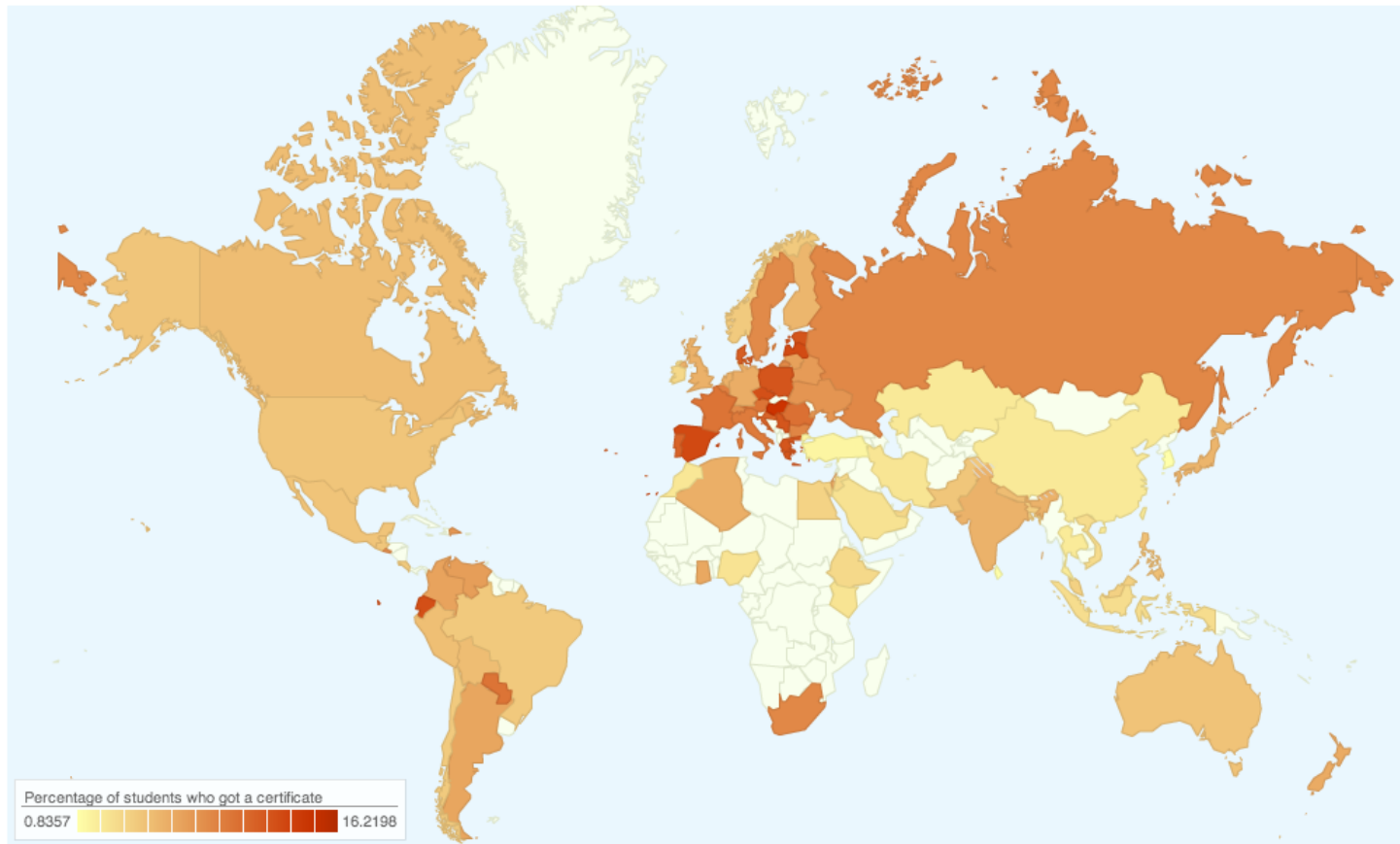
Percentage of students who got a certificate

Fig. 3. Percentage of registrees who received a certificate by country. Interestingly, Eastern Europe had the largest percentage of students who successfully completed 6.002x. Hungary led with 16.2% of students receiving a certificate. The U.S. had 5.1% completion rate.
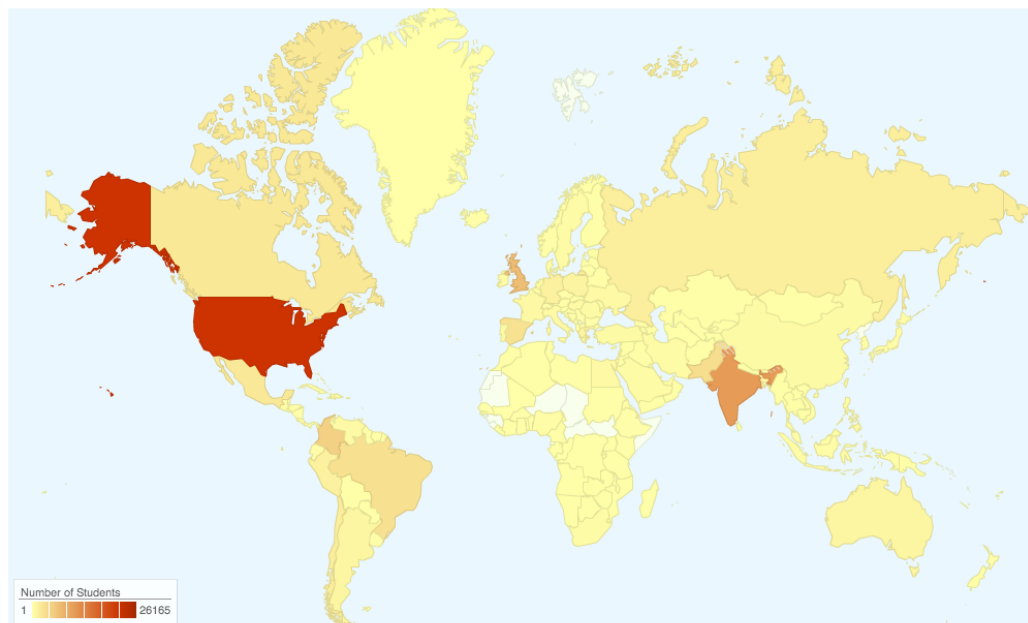
3   Registrations

Fig. 4. The number of students who registered for 6.002x by country. The U.S. and India had the largest absolute number of students.

Percentage of country inhabitants registered to the course: (countries with more than 50,000 inhabitants and IM excluded as IM was 0.0213)
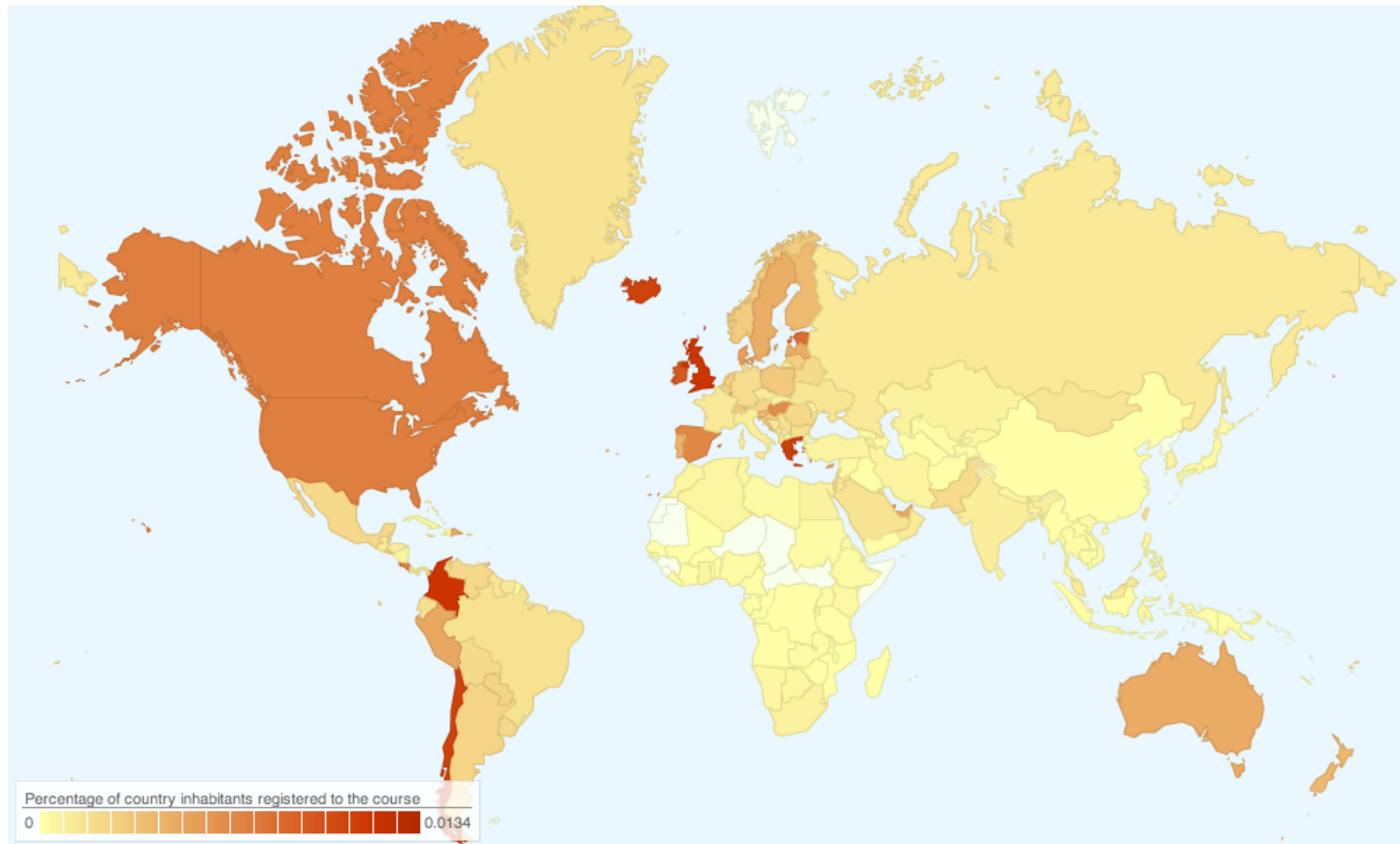

Fig. 5. Percentage of students who registered for 6.002x of the total population of the country. Countries with populations under 50,000 were excluded from the list. Great Britain and Columbia had among the highest percentages with 0.0132% and 0.0134% respectively.
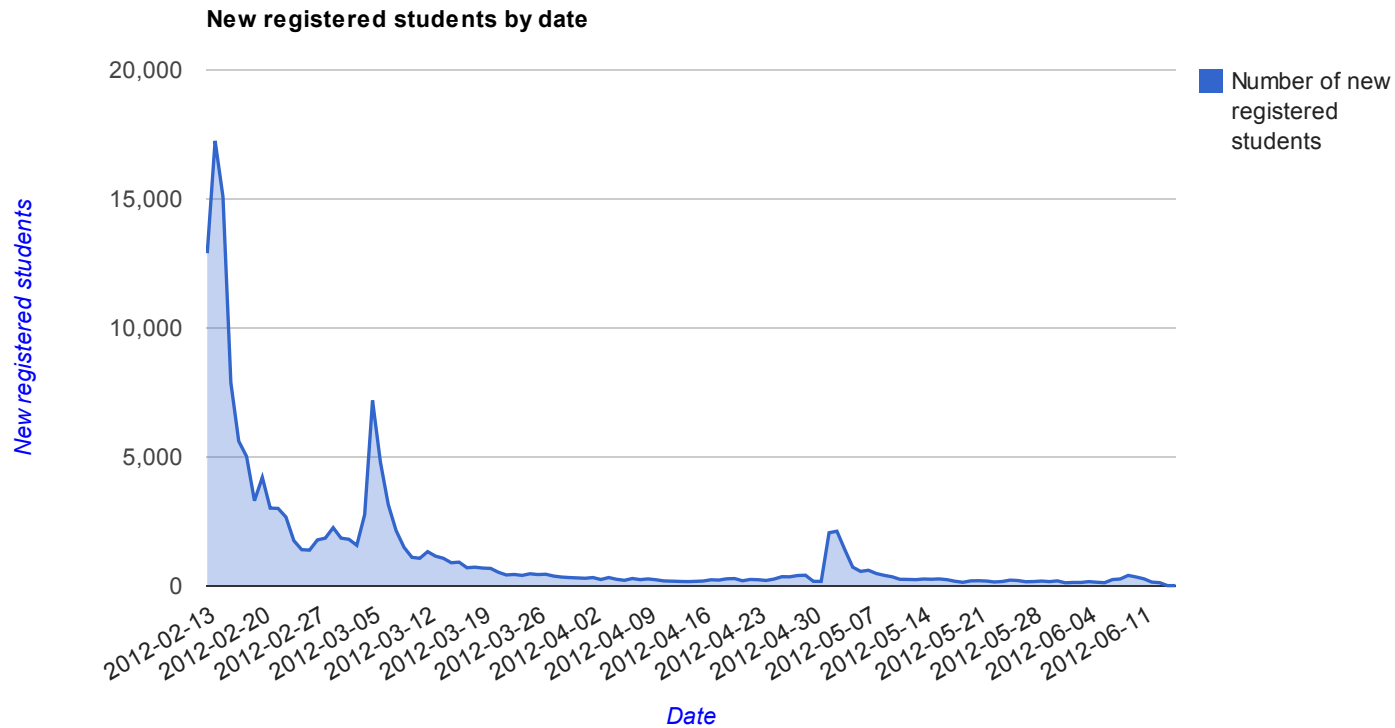
New registered students by date

Fig. 6. The number of students who registered each day since registration opened. Days are delineated by midnight UTC.
From this graph, we see three clear spikes of when users registered for 6.002x. The first and largest, on 02/14/2012, shows the largest number of students registered right after registration opened. 17248 students, about 11% of the students, registered on that date alone. The next spike, around 03/05/2012, corresponds to the start of the course. The final, and smallest spike, around 05/02/2012, corresponds to a series of press releases announcing that Harvard will join MITx to create edX.
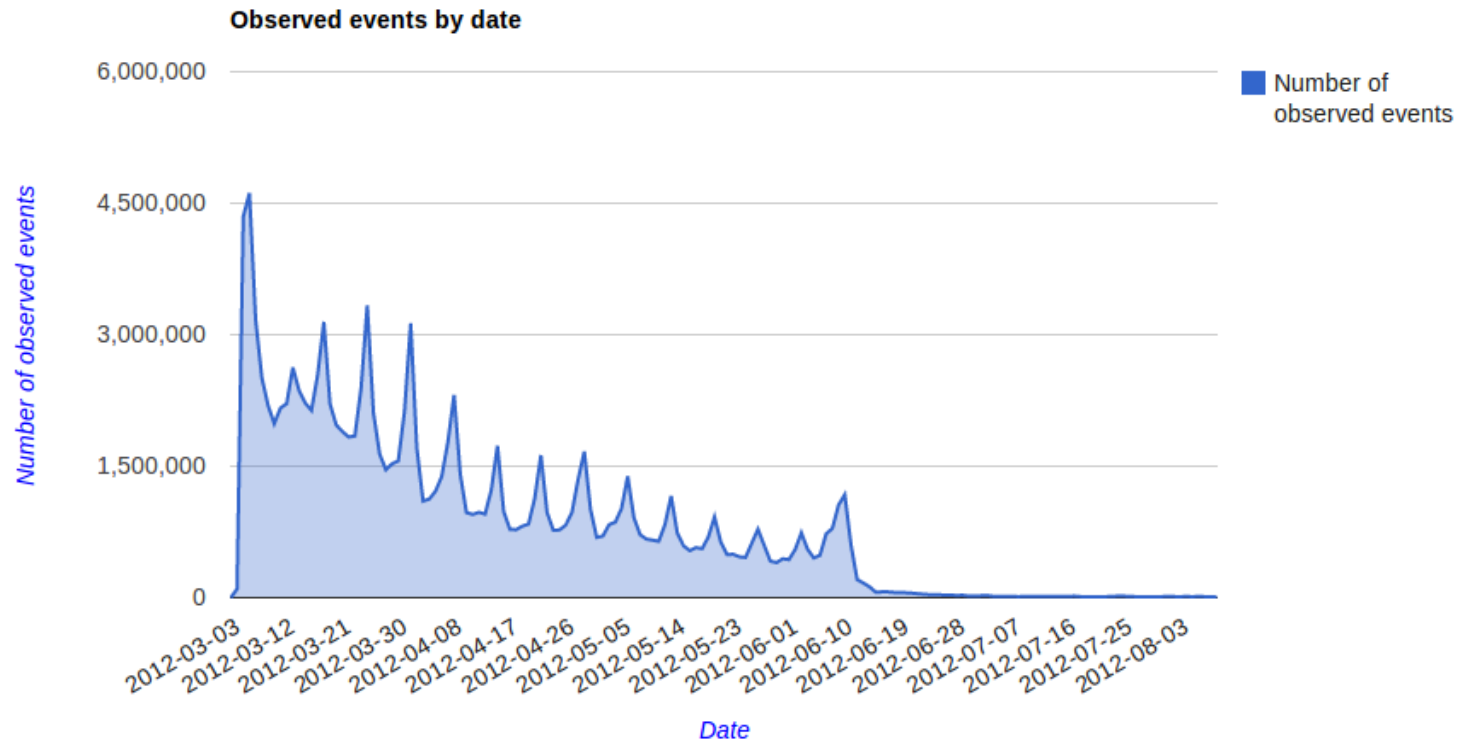
4 Observed Events

Fig. 7. The number of observed events each day of the course Here, an observed event is a user accessing a 6.002x resource. A resource is any page of 6.002x content (including videos, forums, problems, etc.), differentiated by URL.

We see a series of sharp spikes, corresponding to the due dates of homeworks, labs, the midterm and the final. We also see a slow decay of resource access, showing the drop-off of students throughout the semester.

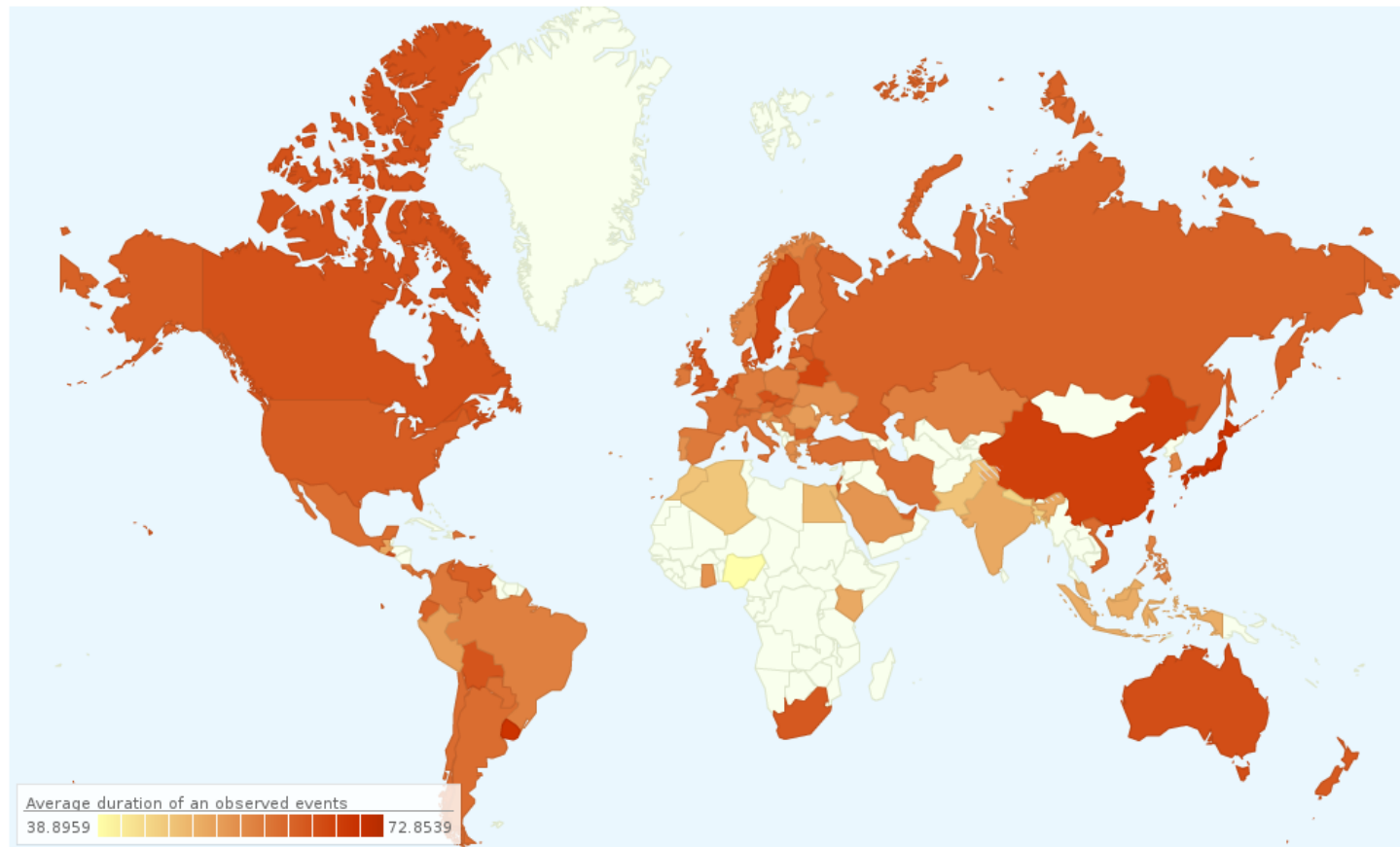Average duration of an observed event (countries with less than 100,000 observed events were removed):

Fig. 8. The duration of observed events, averaged by country. The duration of a resource is difference in timestamps between subsequent resource accesses (per user), with a maximum of 1 hour. A longer average duration would indicate students spending more time on each resource before moving on to another page of the site.

The average duration of observed events is fairly consistent across countries, ranging from around 39 to 73 seconds.
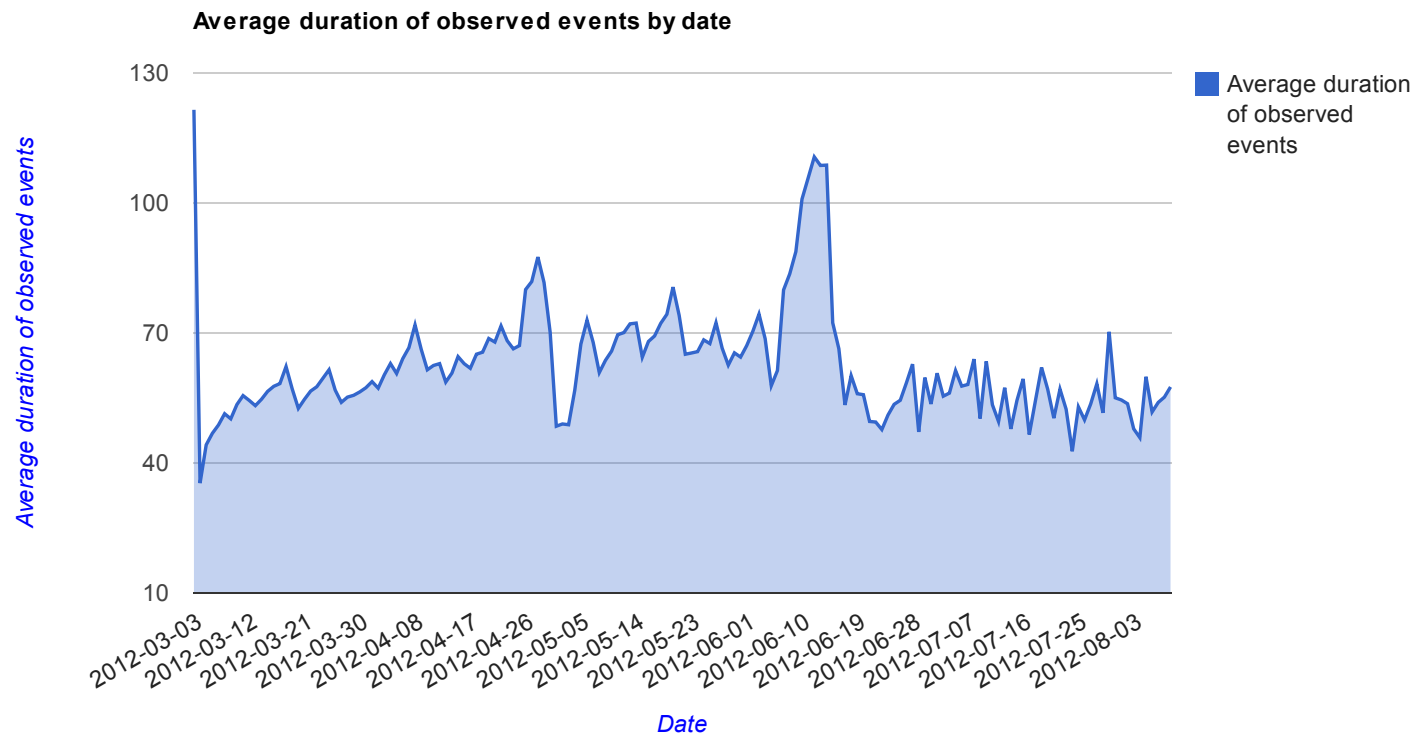
Average duration of observed events by date

Fig. 9. The duration of observed events, averaged by date. There is one noticeable spike around 06/12/2012, after the final exam. This may be due to many students accessing few resources (such as checking a grade or forum). These students may not subsequently view other resources, which could lead to many maximum duration events.
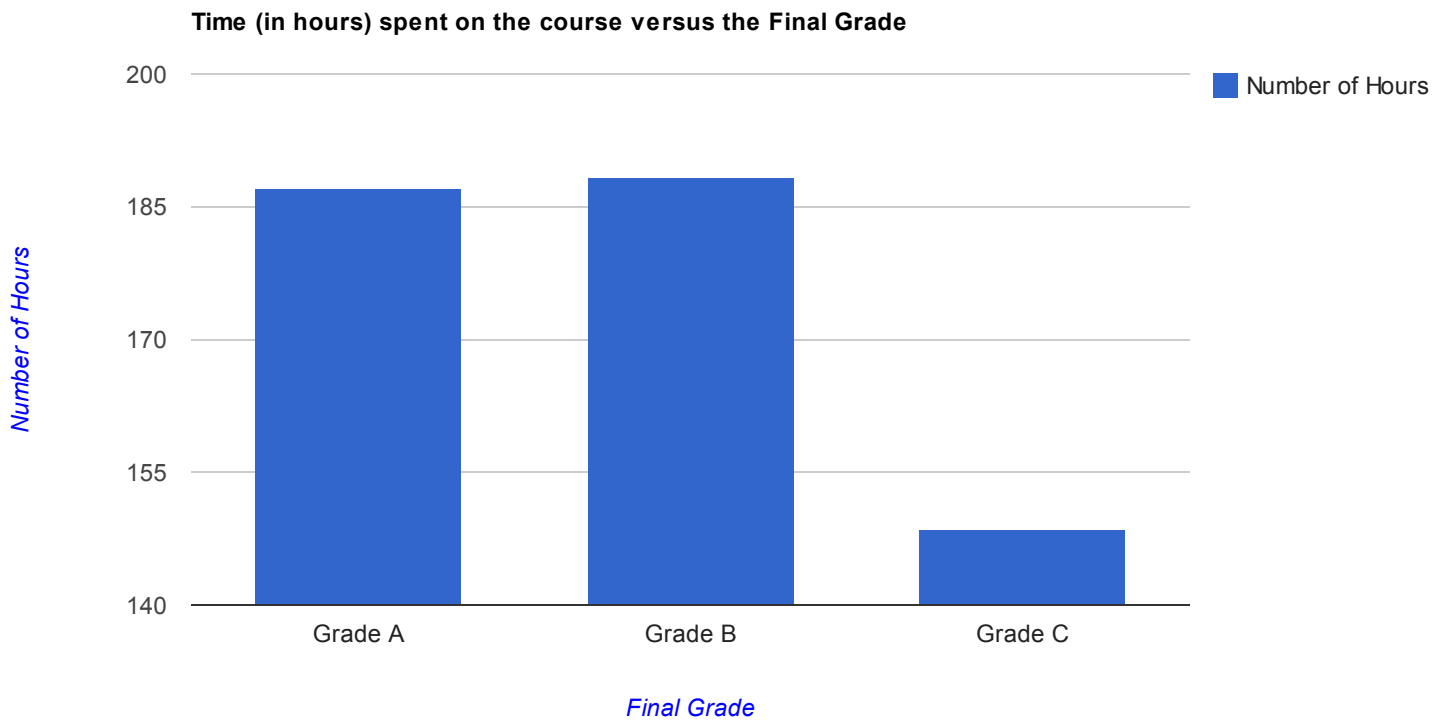
Time (in hours) spent on the course versus the Final Grade

Fig. 10. The cumulative time spent on resources, averaged by grade in the class. It shows a correlation between time spent and score, although A and B grades had relatively similar cumulative times.

5 Submissions

Average number of submissions per country for students who got a certificate (countries with less than 50 students were removed)
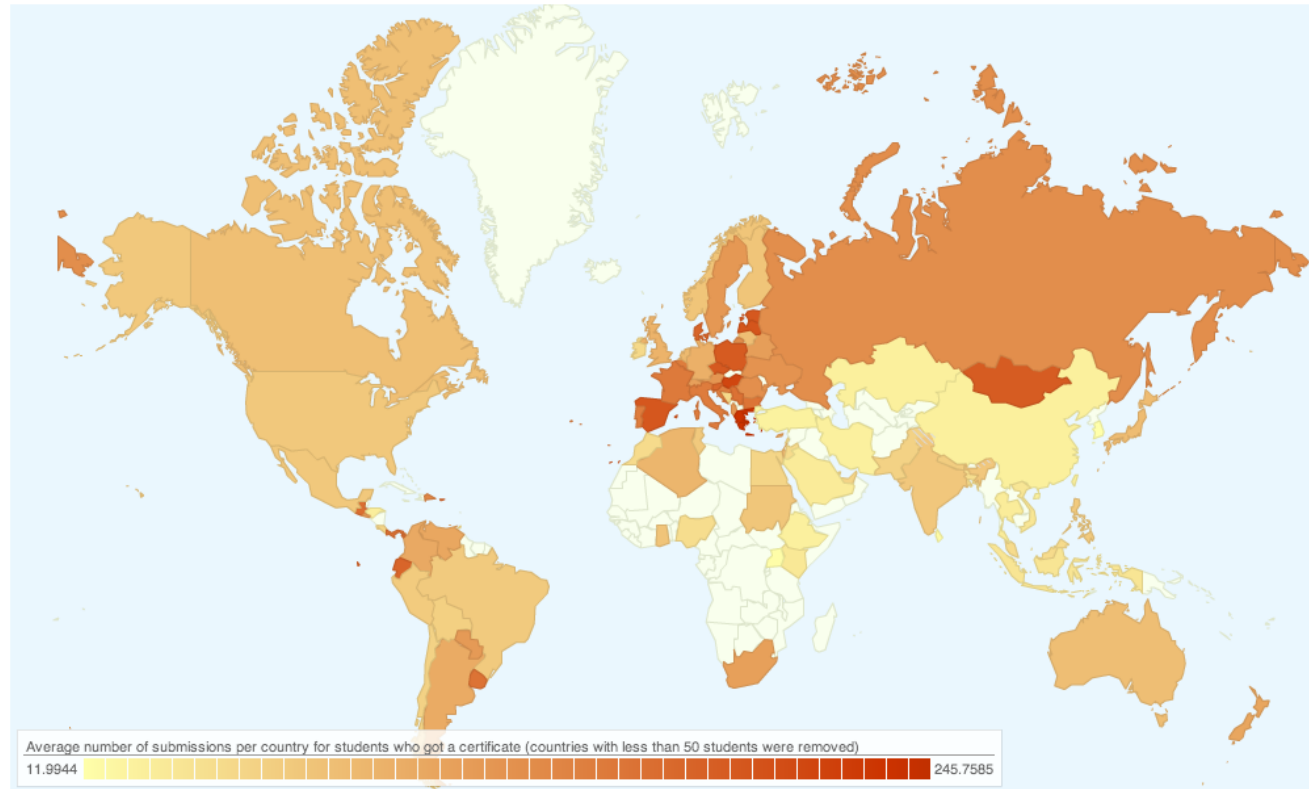
Fig. 11. Average number of submissions per country of students who received a certificate. Students in countries with a larger average submission number show less reluctance in submitting many times before getting the correct answer. Eastern Europe had the largest numbers of average submissions. We see a correlation between this and the percentage of the students who earned a certificate.

Number observed events per submission for students who got a certificate by country having at least 50 students
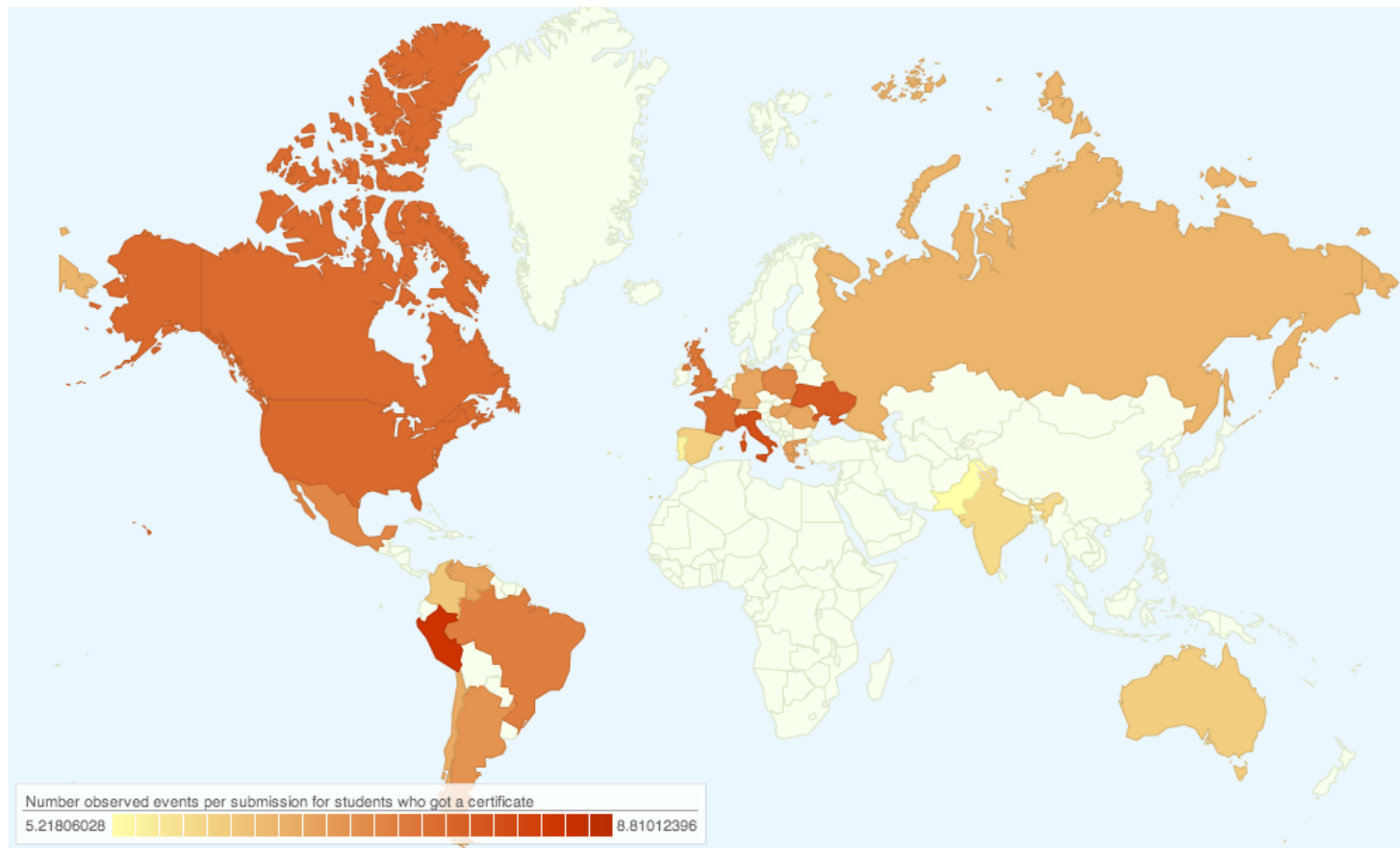

Fig. 12. Observed events per submission per country of students who received a certificate. This number gives a measure of how long a student will spend on course material before submitting an assignment
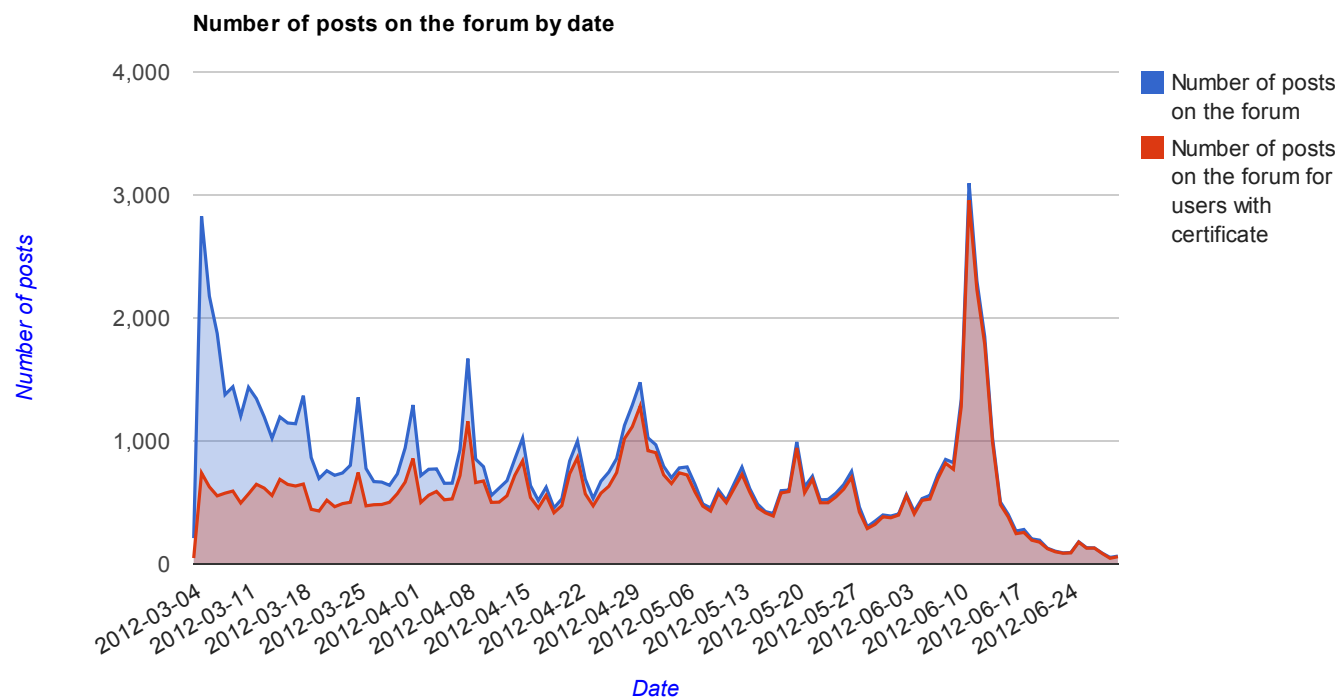
13

6 Forum and Wiki

Fig. 13. The number of forum posts each day after the forum opened. Days are delineated by midnight UTC. Posts by users who earned a certificate are overlaid in red.

In this graph, we see a large spike when the forum first opened. The largest spike in forum posts is on 06/11/2012, one day after the final exam. This most likely is due to students discussing the final and grades on the forum. We also see weekly cycle spikes in forum activity that match the due dates of the homeworks, labs and the midterm.
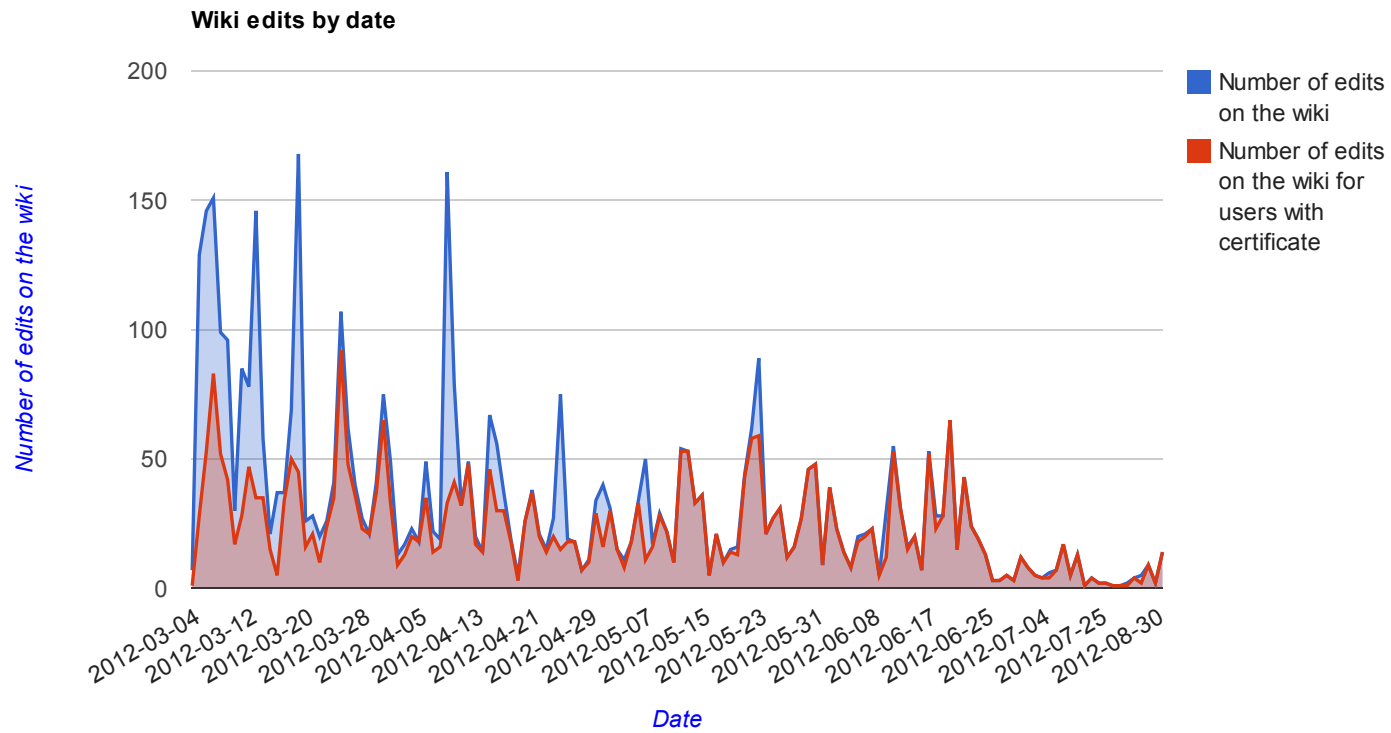
Wiki edits by date

Fig. 14. The number of wiki edits each day after the wiki opened. Days are delineated by midnight UTC. Edits by users who earned a certificate are overlaid in red.

Similarly to the forum posts, we see spikes corresponding to the due dates of homeworks and labs. We do not see much more activity on the wiki around the final or midterm.
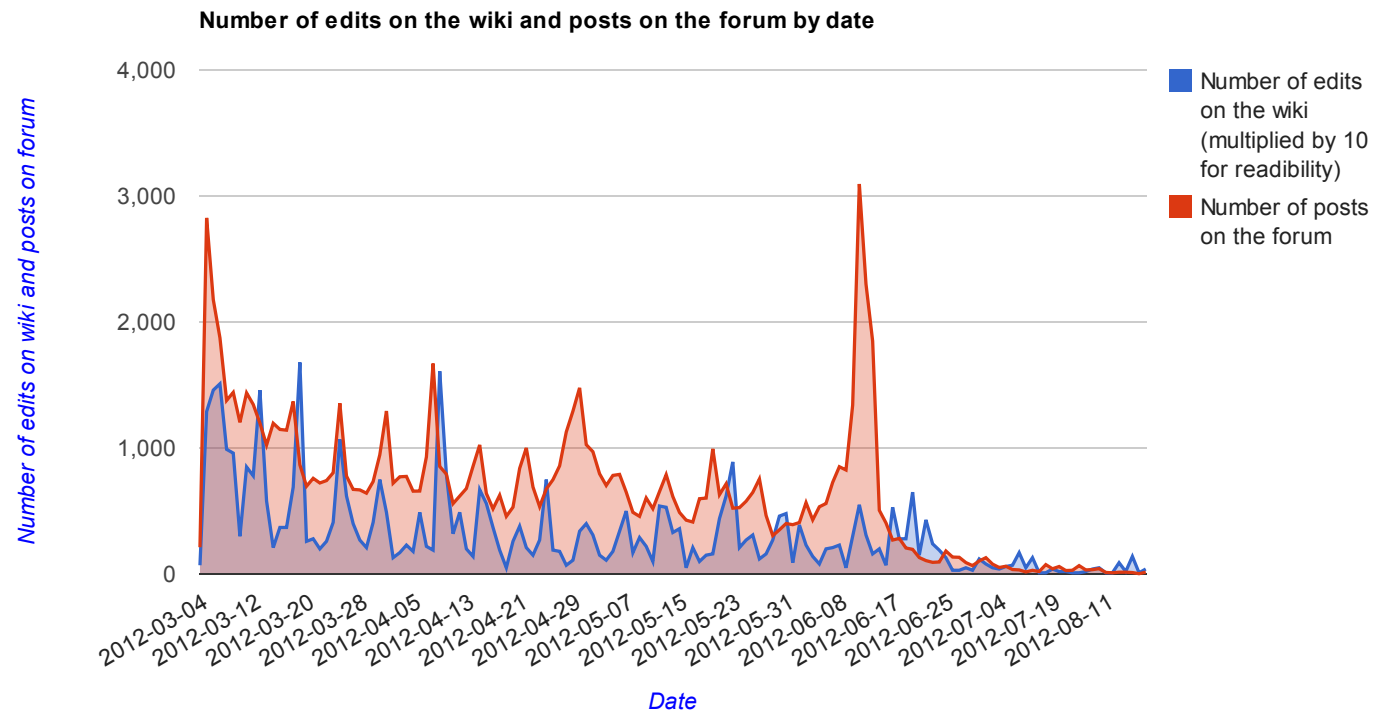
Fig. 15. An overlay of the number of forum posts the wiki edits. Interestingly, we see a 1-3 day delay of the spikes in activity between the two types of collaboration.
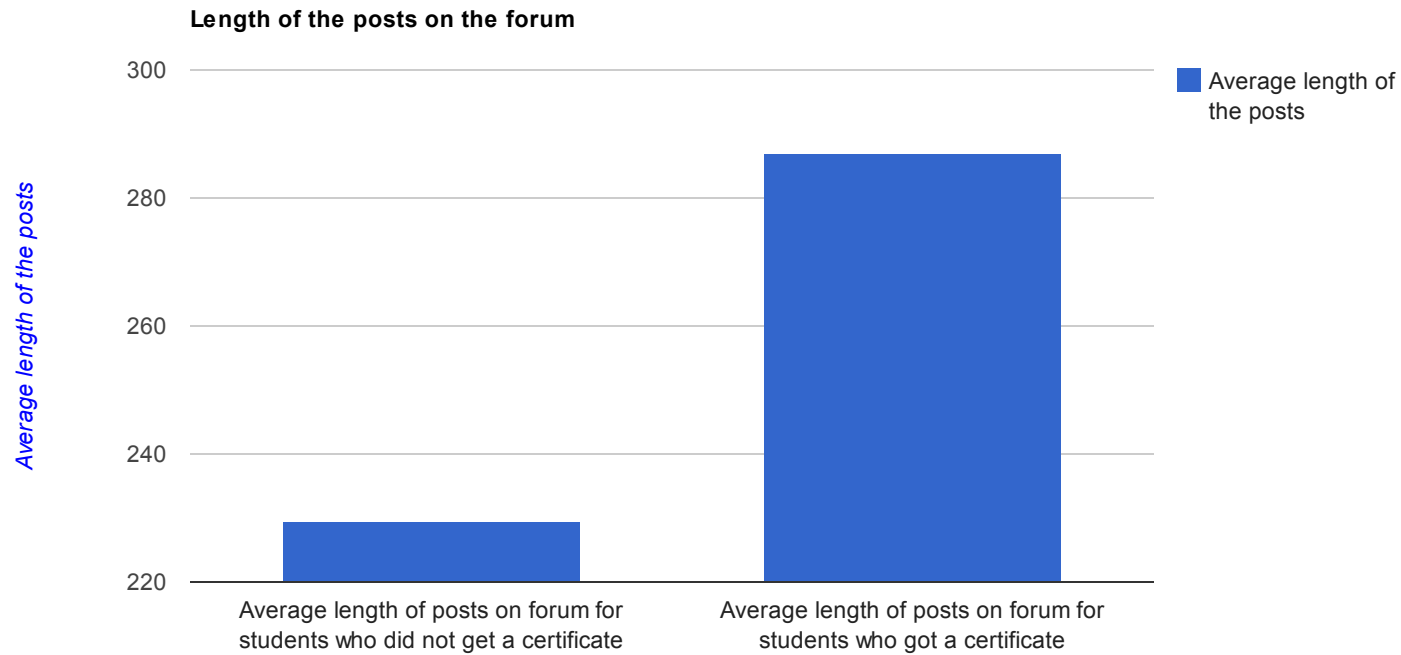
Fig. 16. Average length of forum posts for those users who earned a certificate and those who did not. If a student never posted, he was ignored in this number. The average forum post length for students who earned a certificate is 20% longer than the average for students who did not.

References

1. Franck Dernoncourt, Kalyan Veeramachaneni, Colin Taylor, and Una-May O'Reilly. Methods and tools for analysis of data from MOOCs: edx 6.002x case study. In *Technical Report, MIT*, 2013.