Evaluating Geographic Data in MOOCs

Sergiy O. Nesterko[*]
HarvardX
sergiy_nesterko@harvard.edu

Svetlana Dotsenko
Project Lever

Qiuyi Hu
Statistics Department
Harvard University

Daniel Seaton
MITx

Justin Reich
HarvardX

Isaac Chuang
MIT

Andrew Ho
Graduate School of Education
Harvard University

Abstract

Massive Open Online Courses (MOOCs) exhibit a remarkable geographic diversity in terms of student population. While some researchers evaluated MOOCs geography within a single class, there is no framework for systematically studying the geographic component of MOOCs. Using the example of student population of 18 courses offered by HarvardX, Harvard's division for online learning, we formalize the process of evaluating the geographic data of MOOCs with regard to enrollment and certificate attainment. We report the absolute counts of students from various countries and relate them to baselines of potential learners such as population of the country and the number of English speakers. Our approach allows to identify clusters of countries with similar properties and opens discussion of country-specific factors influencing key metrics of MOOCs. Our findings are relevant for developers and marketers, as well as for future research involving MOOC geography.

1 Introduction

Massive Open Online Courses (MOOCs) have a global reach. MOOC student population exhibits a remarkable diversity in terms of location, gender, demographics, and educational attainment level. In terms of geography, MOOC students came from 194 countries as recorded for an MITx MOOC [3] as well as Coursera courses [18, 14, 19, 2]. In terms of gender, current research suggests that most MOOC students are male (for example, 88% males enrolled in a computer science course studied by [3]; 80.2% in a course taught by [18].

The motivations of students taking MOOCs are also quite varied. The reasons for enrolling into a MOOC class vary from obtaining the knowledge and skills gained as a result from taking the course, to personal challenge, employment/job advancement opportunities, to even the entertainment value of the course and social understanding and friends gained as a result of taking the course [3]. MOOCs attract traditional as well as non-traditional learners [11]. A study by [10] indentifies four clusters of learners with varying learning goals (Auditing, Completing, Disengaging, Sampling).

While there have been some studies of geography and diversity of MOOC students within a class, few, if any authors examined MOOC student demographics on a scale larger than a single class. Most importantly, the geography of MOOCs has received little systematic treatment.

In this paper, we develop a systematic approach to the geographic component of MOOCs, which would prove useful to course developers, marketers, and educational researchers alike. We consider

[*] nesterko.com

baseline populations consisting of learners who potentially can convert into metrics of interest such as register for an online course or earn a certificate. Since the exact composition of baseline populations is an open research question, we suggest and study proxies for them. For the registration metric, we define two proxies – the population of a country [16] and the number of English speakers in a country [5, 7, 13, 6, 8, 9, 1, 15, 4, 17]. For certificate earners and gender composition metrics, we consider the country's number of registrants as baseline population.

Our approach to the geography of MOOCs allows to identify meaningful clusters and patterns for countries with variable characteristics of students and start assessing potential for course adaptation and differential marketing across countries. Our findings motivate the need for further investigation of the apparent geographic factors influencing the effectiveness of MOOCs by country. We develop tools for detecting underserved user populations and highlight opportunities for expanding access and personalizing educational content.

2 Registration

Figure 1 shows the data on all-time student registration for all HarvardX courses (as of September 8, 2013, there are 5 completed and 13 ongoing courses). Table 1 shows data on the top 10 enrolled countries, collectively accounting for over 2/3 of HarvardX enrollment.

As of September 8, 2013, there were 572,899 all-time registrants from 206 countries in 18 HarvardX courses (see Figure 1). Of those, 53% of registrants did not provide address information or a country of origin that our parser could recognize. Using the standard assumption of Missing at Random [12] for the missing data in the absence of evidence to the contrary, we perform the analysis of course registration and country of origin of HarvardX MOOCs.

Several key findings emerge: 1) USA has the most registrants, followed by India and Canada; all five continents are represented among top 10 countries of origin; 2) MOOC's registration varies across countries as measured by the fraction of registrants to the country's population; 3) countries with large proportion of English speakers in their population tend to be over-represented among top 10 countries of origin of HarvardX MOOC registrants; 4) male registrants represent a majority in most countries of the world.

1. Most Registrants from US, India, Canada; All 5 continents represented in Top 10

The first variable for measuring MOOCs' geography is the number of registrants and their geographic distribution. This variable is a proxy for the course's popularity worldwide, and presumably depends on the level of student interest, accessibility, and/or course marketing. In our analysis, we include all students that indicated their interest in a class by registering on HarvardX platform; students registering for more than 1 course are counted once for each class.

In our analysis, students from the US account for over 242,000, of 42.3% of all Harvardx MOOC registrants. The USA is followed by India (over 54,000 students and 9.47%) and Canada (less than 22,000 students and 3.81%).

Countries from all continents (with the exception of Antarctica) are represented among Top 10 countries of origin for the total number of HarvardX MOOC registrants. Australia holds the 4th place with over 12,000 students, which represents 2.18% of all MOOC registrants. In Africa, Nigeria is the most enrolled country with the estimated 12,067 students (2.11% of worldwide registrants). Next, in South America, Brazil has 11,243 students (1.96%) who registered for HarvardX courses. Next, Spain has the most students in Europe with 10,582 (1.85%); the second largest enrolled European country that also is on Top 10 list is Great Britain with 8,066 students. Finally, Philippines and Pakistan are countries of origin for almost 20,000 students.

2. MOOC usage as registration to population varies significantly across countries

The high-level registration data on Figure 1 do not control for the countries' populations. Indeed, if HarvardX online courses were equally used in all countries around the world, one would expect registration numbers to be a constant proportion of the country's population. Hence, we propose a variable that would account for the number of registrants relative to the country's population. This variable can show not only how many students registered for a MOOC in a given country, but also how many students could be potentially interested or reached with more marketing.

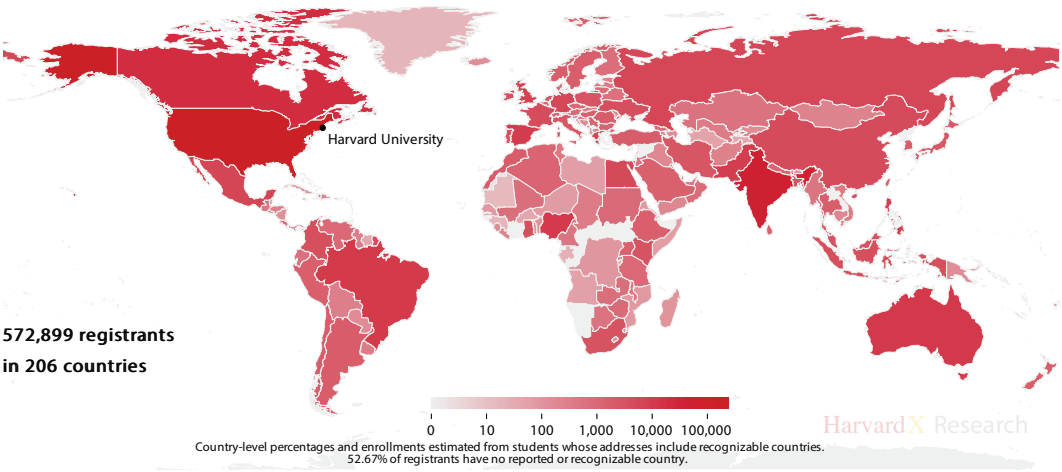

Figure 1: Estimated registration for HarvardX courses (all-time) by country as of September 8, 2013.

Table 1: Estimated enrollment and worldwide proportions of registrants in Top 10 countries of origin for registrants for 18 HarvardX courses as of September 8, 2013, sorted most to least registered; country registration percentage as the percentage of all HarvardX students that registered for HarvardX MOOC; country registrants as a fraction of the country's population, in $\frac{1}{100\%}$; country registrants as a fraction of English speakers, in $\frac{1}{100\%}$.

Continent	Country	Registration	Reg. %	Reg. / Pop., $\frac{1}{100}$%	R. / Eng. P., $\frac{1}{100}$%
N.America	United States	242,279	42.29%	7.64	9.06
Asia	India	54,230	9.47%	0.44	4.33
N. America	Canada	21,853	3.81%	6.22	8.66
Australia	Australia	12,474	2.18%	5.38	7.19
Africa	Nigeria	12,067	2.11%	0.70	1.53
S. America	Brazil	11,243	1.96%	0.56	11.83
Europe	Spain	10,582	1.85%	2.27	10.75
Asia	Philippines	10,099	1.76%	1.03	1.33
Asia	Pakistan	9,505	1.66%	0.52	1.07
Europe	United Kingdom	8,066	1.41%	1.27	1.35

For example, HarvardX registrants in the United States account for 7.64 one-hundredths of a percent (basis points) of the country's population. For India the number is just 0.44 basis points - although India is the second most-enrolled country, HarvardX MOOCs are least used in India among the top ten enrolled countries. China's registrants account for only 0.04 basis points. In Spain and Greece, the fractions of registrants to country population are high: 2.27 basis points in Spain and 6.96 basis points in Greece.

Some of potential explanations for such variety include but are not limited to the proliferation of English language, differences in marketing, cultural environment, internet access, perceived value of education, interest in the US education, state of the economy, among others. Future studies will need to determine the importance of various factors for MOOC registration. Following is an example of accounting for one such factor, number of English speakers in a country.

3. English-speaking countries are over-represented among Top 10

To account for one of the potential factors cited above (number of English speakers), we examined the relationship between the registration rate and the number of English speakers in a country.

Among top 10 countries of origin, six have over 50% of English speakers in their population. However, the other four countries have a significant number of English speakers as well, such as India (125M English speakers, accounting for 9.47% of registrants). Nigeria has about 79M English

speakers; Pakistan about 89M. Another deviation is Brazil (9.5M English speakers, or 5% of its population).

4. Male registrants represent a majority in all but 10 countries of the world

Within all HarvardX courses, we estimate 63.4% males and 36.6% females using Missing at Random assumption for missing or undisclosed gender data with 8.4% missing data. Gender composition by country varies (see Figure 2).

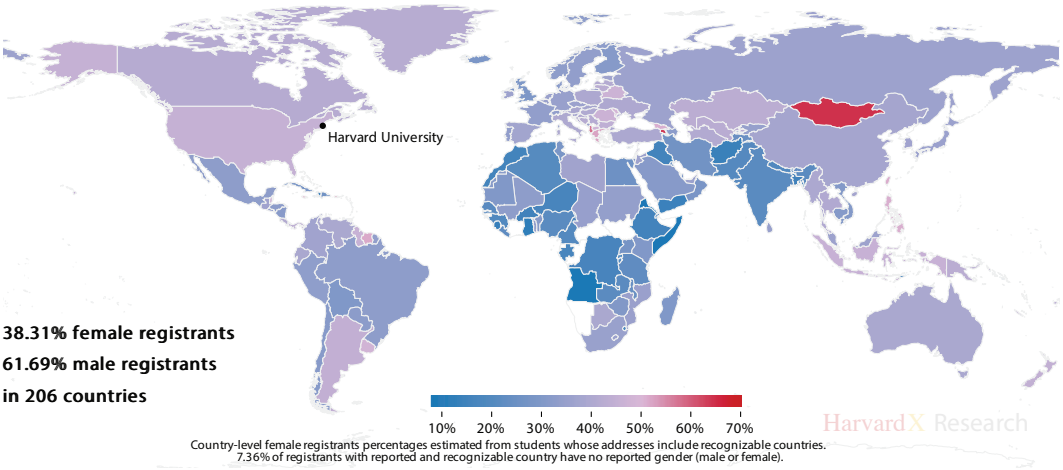

Figure 2: Estimated proportion of female registrants for HarvardX courses by country as of September 8, 2013.

Some individual courses (such as SPU27x, The Science of Cooking, launching October 2013), register relatively equal gender percentages of registrants (50.71% male and 49.29% female). Worldwide, there are only 10 countries with more than 50% female registrants. Among them, Philippines has the largest estimated number of registrants, just over 10,000 (1.78% of worldwide enrollment), with estimated 50.9% females. The second country with most female registrants is Greece with estimated 7,522 registrants and 52.3% females.

3 Certificate attainment

HarvardX allows students who fulfilled course requirements and completed a course to obtain certificates of completion. As of September 8, there were 15,895 certificates of completion issued to students in 155 countries by 5 completed HarvardX MOOCs (see Figure 3). All numbers of certificates are estimated. Estimation is based on the Missing at Random assumption with 46% of registrants missing address information. Table 2 presents top 10 countries by number of issued certificates, together accounting for almost 2/3 of all certificates issued by HarvardX to date (64.1%).

Several findings emerge from certificate completion data: 1) most countries that account for the most registrants also have the most certificate earners; 2) if ranked by percentage of registrants who pursued and received certificates, rather than absolute number of certificates, European countries lead strongly; 3) international students, on average, are more more likely to receive a certificate than their American counterparts.

1. Most certificates in countries with highest registration numbers

Similarly to the top countries by origin of registrants, United States and India account for most certificate earners, with 4,674 certificates (29.41% of total certificates issued), and 1,606 certificates (10.1% of all certificates), respectively. However, some of the other Top 10 countries changed.

The third country with most certificates is Spain (725 certificates, 4.56% of all), which was also on the top 10 registrants' list. Spain's European counterpart Greece, which was not represented on the top 10 countries of origin for registrants, holds the 4th place in terms of total number of certificates

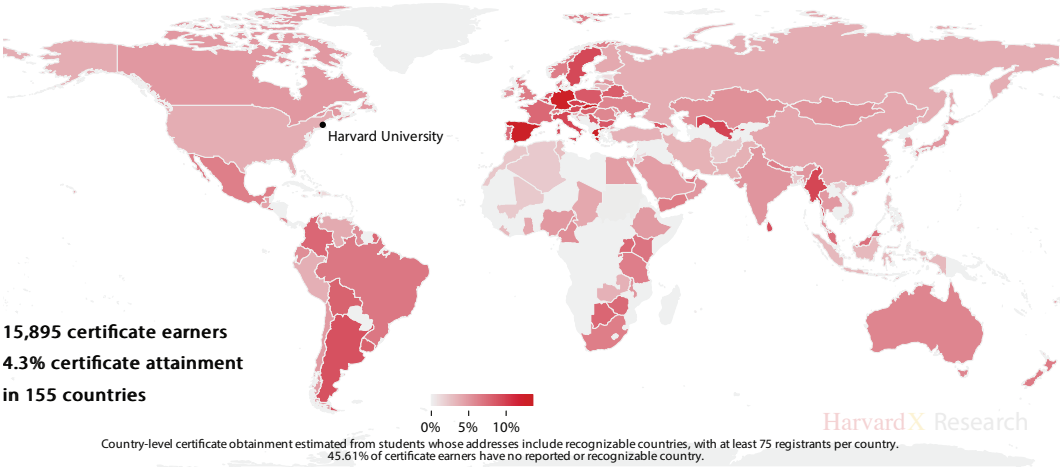

Figure 3: Estimated certificate attainment rates for 5 completed HarvardX MOOCs by country as of September 8, 2013. Estimation performed for countries with at least 75 registrants.

Table 2: Estimated certificate earner numbers, certificate attainment rate relative to number of registrants, and certificate earner proportion in worldwide total for top 10 countries by number of certificate earners for 5 completed HarvardX MOOCs.

Continent	Country	Certificates	Cert./Reg.	Certificates %
N. America	United States	4,674	3.72%	29.41%
Asia	India	1,606	5.12%	10.1%
Europe	Spain	725	12.72%	4.56%
Europe	Greece	623	13.58%	3.92%
N. America	Canada	567	4.88%	3.57%
Europe	Germany	455	12.68%	2.86%
S. America	Brazil	439	6.82%	2.76%
Australia	Australia	411	6.05%	2.59%
Africa	Nigeria	390	4.94%	2.45%
Europe	United Kingdom	299	6.53%	1.88%

(623 certificates, 3.92% of total). Another notable addition to the Top 10 list is Germany, holding the 6th place with 455 certificates and 2.86% of total awarded certificates.

2. European countries have the highest certificate attainment rate, not absolute number

In addition to considering the total number of certificates, we analyzed the rate of certificate attainment – the number of certificates divided by the number of registrants in a given country. A high certificate attainment rate can be a strong manifestation of academic persistence or proficiency. Alternatively, it may indicate weaker marketing in a given country so that more motivated students self-select into registering and completing. In case maximizing the number of certificate earners is an objective of a MOOC provider, it can be useful to study the geography of certificate attainment.

The mean rate of certificate attainment world wide for MOOCs considered in our analysis is 4.3%. The country with the highest certificate attainment worldwide is Greece with estimated 13.6% of initial registrants completing the course with a certificate. Students from European countries display high persistence with some of the highest certification rates worldwide. Five of the six countries with the highest certicifate attainment rates in the world are European: Greece (13.6%), Spain (13.1%), Slovakia (13.0%), Germany (12.8%), and Czech Republic (10.1%).

Latin American countries with the highest enrollment tend to have higher than average certificate attainment rates of 6-9%. For example, certificate attainment is estimated at 6.82% for Brazilian students, 7.69% for students in Colombia, and 9.01% for those in Argentina. An exception is Peru with 3.58% certificate attainment rate. Asian countries exhibit variable certificate attainment rates.

Three countries with the highest estimated values are Uzbekistan (9.77%), Myanmar (9.73%), and Sri Lanka (9.23%). The Four Asian Tigers (Hong Kong, Singapore, South Korea, and Taiwan) all have higher than average certificate attainment rates (5.93%, 6.71%, 4.59%, 4.73%, respectively). In Africa, students in countries along the east coast form a coherent cluster with 7.8% certificate attainment in Botswana, 7% in Kenya, and 6.4% in South Africa. The highest-enrolled country Nigeria (estimated 9,445 students) accounts for 4.94% certification rate.

3. International students receive certificates at higher rate than Americans

While the the number of both course registrants and certificate earners is the highest in the United States, the proportions of registrants and certificate earners in the worldwide student body are different. US accounts for estimated 42% of worldwide registrants (see Table 1) but only for 29% of certificate earners. The rate of certificate attainment for the US is 3.72%, which is lower than worldwide average of 4.3%. This suggests that international students, on average, are more persistent in their studies then their American counterparts taking HarvardX online courses.

Understanding the reasons of the apparent difference in persistence across student populations from different countries will help develop online learning experiences accessible and useful to more students worldwide.

4 Discussion

In this study, we develop an approach to analyzing the geographic components of MOOCs that considers fractions of realized registrants or certificate earners to populations of potential learners in each country. The framework allows to uncover meaningful geographic variation for HarvardX MOOCs and generalizes to other course providers.

We suggest several directions for future research. The proxies for baseline populations used in our study (i.e. country population size, number of English speakers, registrants) can be refined, tested, and supplemented by others such as the size of middle class, number of people with Internet access, number of high school graduates, and others. Future research is also needed on country-specific factors influencing key MOOC metrics such as registration and certificate attainment. Such analyses will help direct efforts of course developers, researchers, and marketers of MOOCs.

References

[1] Australian Bureau of Statistics. Population Clock, 2013. URL http://www.abs.gov.au/ausstats/abs@.nsf/94713ad445ff1425ca25682000192af2/1647509ef7e25faaca2568a900154b63?OpenDocument.

[2] Tucker Balch. MOOC Student Demographics (Spring 2013), 2013. URL http://augmentedtrader.wordpress.com/2013/01/27/mooc-student-demographics/.

[3] Lori Breslow, David Pritchard, Jennifer DeBoer, Glenda Stump, Andrew Ho, and Daniel Seaton. Studying Learning in the Worldwide Classroom: Research into edX's First MOOC. *Research & Practice in Assessment*, 8:13–25, 2013. URL http://www.rpajournal.com/dev/wp-content/uploads/2013/05/SF2.pdf.

[4] Census of India. Office of the registrar general & census commissioner.

[5] David Crystal. *English as a global language*. Cambridge University Press, 2003. URL http://catdir.loc.gov/catdir/samples/cam041/2003282119.pdf.

[6] Eurobarometer. Eurobarometer, 2012.

[7] Euromonitor International. Pakistan Government Statistics Division, 2013. URL http://www.teachingenglish.org.uk/sites/teacheng/files/Euromonitor%20Report%20A4.pdf.

[8] IBOPE. Target Group Index, 2013.

[9] Kelechukwu Uchechukwu Ihemere. A Basic Description and Analytic Treatment of Noun Clauses in Nigerian Pidgin. *Nordic Journal of African Studies*, 15(3):296–313, 2006.

[10] Rene Kizilcec, Chris Piech, and Emily Schneider. Deconstructing disengagement: analyzing learner subpopulations in massive open online courses, 2013. URL http://dl.acm.org/citation.cfm?doid=2460296.2460330.

[11] Alanna Klapp. MOOCs Open Doors for Diverse Student Body. *Diversity Journal*, 2013. URL http://www.diversityjournal.com/10107-moocs-open-doors-for-diverse-student-body/.

[12] Roderick J A Little and Donald B Rubin. Statistical analysis with missing data. 2002.

[13] Republic of the Philippines, National Statistics Office, 2000. URL http://www.census.gov.ph/data/sectordata/sr05153tx.html.

[14] Charles Severance. Visualizing the Geographic Distribution of my Coursera Course, 2012. URL http://www.dr-chuck.com/csev-blog/2012/09/geographic-distribution-of-my-coursera-course/.

[15] Statistics Canada, 2013. URL http://www.statcan.gc.ca/start-debut-eng.html.

[16] UNDESA. Population, total both sexes (thousands), 2011. URL http://hdr.undp.org.

[17] United States Census Bureau. Language Use in the United States, 2007. URL http://www.census.gov/hhes/socdemo/language/data/acs/ACS-12.pdf.

[18] Tim van Boxtel. Interesting Participant Demographics from a Coursera course, 2013. URL http://tim.vanboxtel.ca/wp/interesting-participant-demographics-from-a-coursera-course/.

[19] James Yager. Massive Open Online Courses with a Global Outreach, 2013. URL http://www.aahcdc.org/Portals/0/mtgs/if13/Yager.Jim_MassiveOpenOnlineCourseswithGlobalOutreach.pdf.