

# A Learning Analytics approach to Improving Course Design for the Global Community

## Introduction

Online learning, particularly through Massive Open Online Courses (MOOCs), is gaining wider use and acceptance as a form of educational advancement. Many MOOCs now offer certification to users who complete the content correctly to a specified percentage. Bayeck et al examined the demographics of MOOC users and the effect of demographics on learner motivation. Both the Massachusetts Institute of Technology (MIT) and Harvard now offer online versions of their undergraduate courses through MITx and Harvardx, respectively. To assess the effectiveness of online learning, MITx and Harvardx jointly gathered data from their online learners over the course of the 2012-2013 school year. Their data, made publicly available after deidentification, contributes to the growing Learning Analytics approach to improving education.

By applying Learning Analytics to the edX dataset, we hope to gain insight that could be used to improve the online learning experience for users from all countries.

## Learning objectives

Our primary learning objective was to learn how students in different countries engage with course material and how we can use that information to inform future course design. We began with the question, *who are our users and how does geographical location affect course interaction?* We divided this research question into two phases: learner *characteristics* and learner *participation*.

### ***Learning Objective 1:***

To investigate learner characteristics, we first looked at the percent of learners within each country who registered, viewed, explored, or became certified for each course. We asked, *what are the demographic characteristics of our users and are there trends across countries?* By comparing age, gender and education level status across different regions, we sought to learn if there are major trends to MOOC use across countries or regions and seeking in next steps to

identify barriers to accessibility.

### ***Learning Objective 2:***

To investigate learner participation, we look at two means through which online learners interact with course material: watching videos and making forum posts. Both of these activities are optional, so we determined these variables to be two metrics of course participation, which could reflect learner interest in the material. We hypothesized that learners who watched more videos and were more active on forum discussions were more engaged in the course material. We sought to see if location has any effect on this participation. We asked, *how does geographical location affect course participation through online videos and forum posts?* We began by analyzing the average number of video clicks and forum posts per learner in each country and compared that to educational index, as well as internet speed in the case of video clicks. Our research question became, *how does educational accessibility, as measured by educational index and internet speed, affect online course participation, as measured through the average number of video clicks and forum posts for each country?*

## User and context

The data generated from this study is intended to be used by course and media developers for edX. Developers could learn both what types of courses pique and retain student interest and final certification, as well as how to engage learners along the journey through videos and forum outlets. Additionally, we analyzed this data by comparing across different countries in order to see if this influences how learners approached courses. This can aid the future development of successful courses that are accessible to users in the global learning community.

*Learning Objective 1 Application:* Analyzing learner interest in the courses could aid future course recommendations to learners based on location. This could also reveal what courses may need improvement to be more accessible in certain locations.

*Learning Objective 2 Application:* We chose comparisons with educational index (for average video clicks and forum posts) and internet speed (for average video clicks) to learn about accessibility issues that might affect course participation. Our hope was to use these insights to improve course design and accessibility for specific regions.

## Design and Method

For the analysis of demographic data, the data was examined for incomplete variables and observations without specified country were removed. Furthermore, Education Indices and Development Status were added to countries and countries were sorted into regions.

Using ggplot in R, analysis of age of learners, gender of learner and the certification rate to education index comparisons were performed with appropriate plotting of data.

For the AVPU analysis, the sample user data in all the courses was chosen to meet the following requirements:

- 1) All the users were certified. This selects for users who completed the course and are guaranteed to have had the intention of completing the course. Thus, users who might have just registered, perused, or fallen behind are filtered out.
- 2) All the users have at least 1 video play. This weakly filters out “cheaters” who have duplicate accounts as this was shown to be a predictor for cheating by Alexandron et al. It also gets rid of users with missing video play data.
- 3) All the users were from named countries. Data is typically available for specific countries and due to the de-identification process, we cannot make assumptions about users from regions such as “Other Africa” or “Unknown.”

Besides the reading through the lecture notes and doing assignments in the courses, there are two other instructional tools that are used in MOOCs for which we wanted to measure and compare their effectiveness. These are the use of videos and the forum. Because each course has different numbers of videos, a further restriction on the data is that any comparisons can only be made within the same course since courses have different numbers and types of videos for which we do not have data. To define quantities that can be used for comparison between different countries within the same course, we define:

**Average Video Plays per User (AVPU)** - This is defined as the total number of video plays in a course per country divided by the total number of users in a course in each country normalized by the largest AVPU so that the scale is from 0 to 1.

**Average number of Forum Posts per User per country (AVFPU)** - This is defined as the total number of forum posts in a course per country divided by the total number of users in a course for each country normalized by the largest AVFPU.

A preliminary analysis was performed to determine if there was any categorical relation between the AVPU and AVFPU and country. Bar graphs were plotted in Fig. 1 and Fig. 2 to compare values between countries. It was noticed that developed countries, particularly in Europe dominated both AVPU and AVFPU. The dominance of developed, except for a few anomalies like Egypt, raised the question of what factors specific to developed countries could be responsible for them having higher AVPU and AVFPU.

Two factors for which we could find data were hypothesized as affecting the AVPU and AVFPU.

**Internet speed** - This could affect how willing users were to play videos repeatedly (AVPU).

This data was collected from the Akamai state of the Internet Report in 2012.

**Education Index** - This could affect how comfortable users were to contribute to forums as well as how many times they played video; for example in non-English speaking countries, users might play videos more times to understand, if the country has a lower education index. This

data was collected from United Nations Development Programme's Human Development Reports. We used the Education Indexes from 2013, since the dates for which data was collected was Fall 2012-Spring 2013. There is no significant difference between Education Indexes reported in 2012 or 2013.

Note that the two factors are not decoupled and there are likely more factors that contribute to the results but to set a baseline, the following assumptions were made:

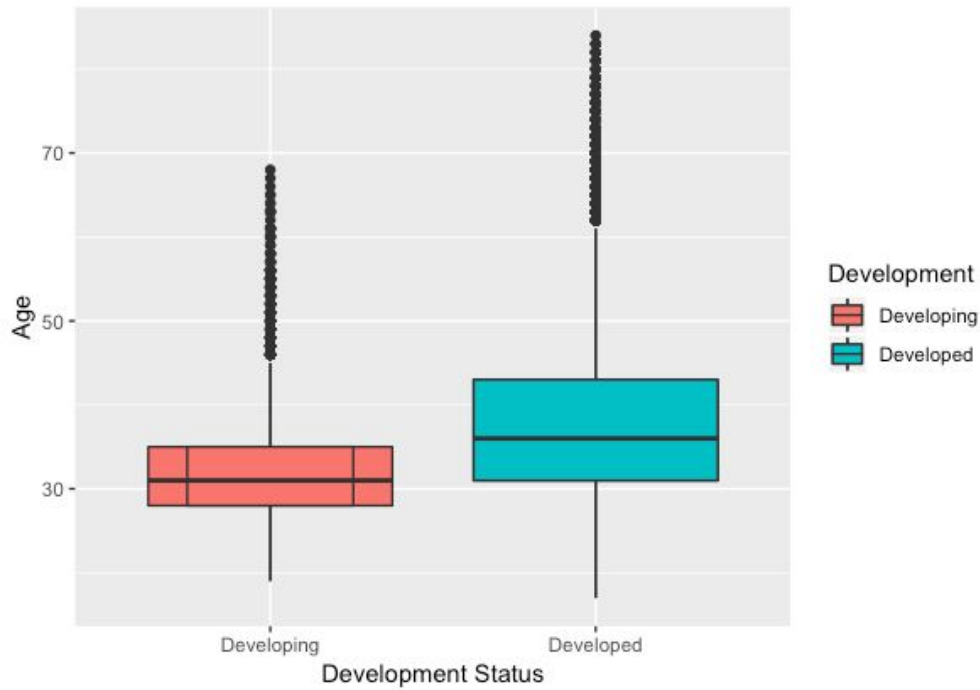
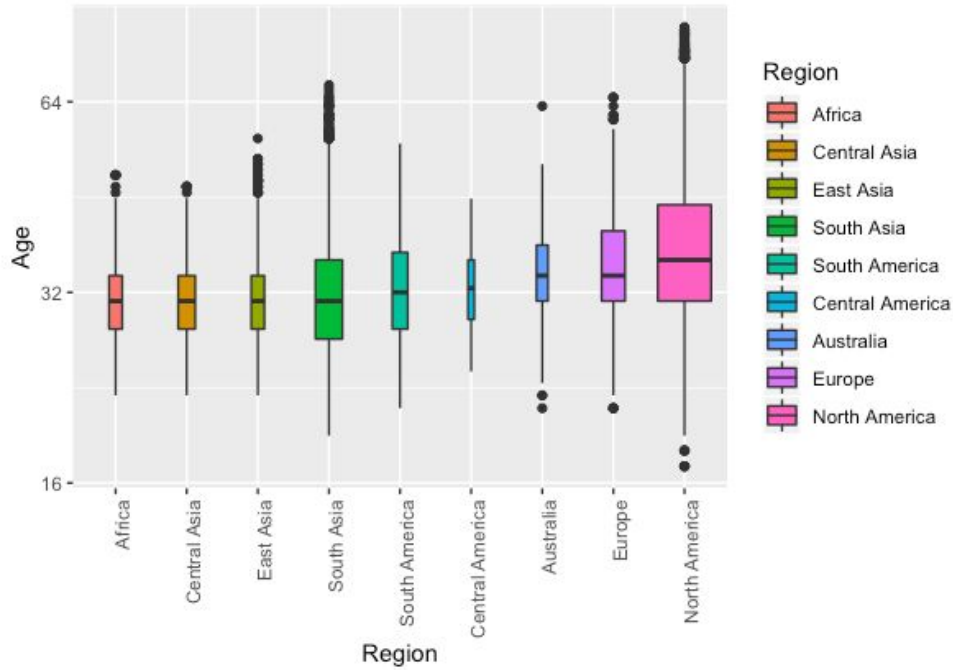
- 1) A video play in the data suggests the user played the entire video
- 2) The users in each country are representative enough that they can be represented by one metric such as the Average Internet Speed and Education Index

The results based on these assumptions can then be used to determine what kind of data would need to be collected in future iterations to improve reliability of results.

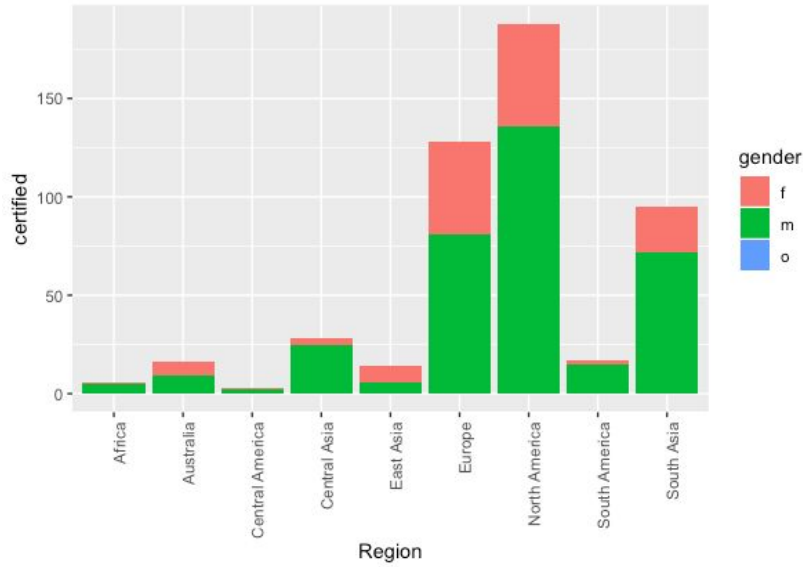
Scatter plots of AVPU against Internet Speed and Education Index were done similar to those in Figure. 5-8 to find any correlation between the two. In addition, scatter plots of AVFPU against Education Index were also done in Fig. 9-12 for a few sample courses

## Results

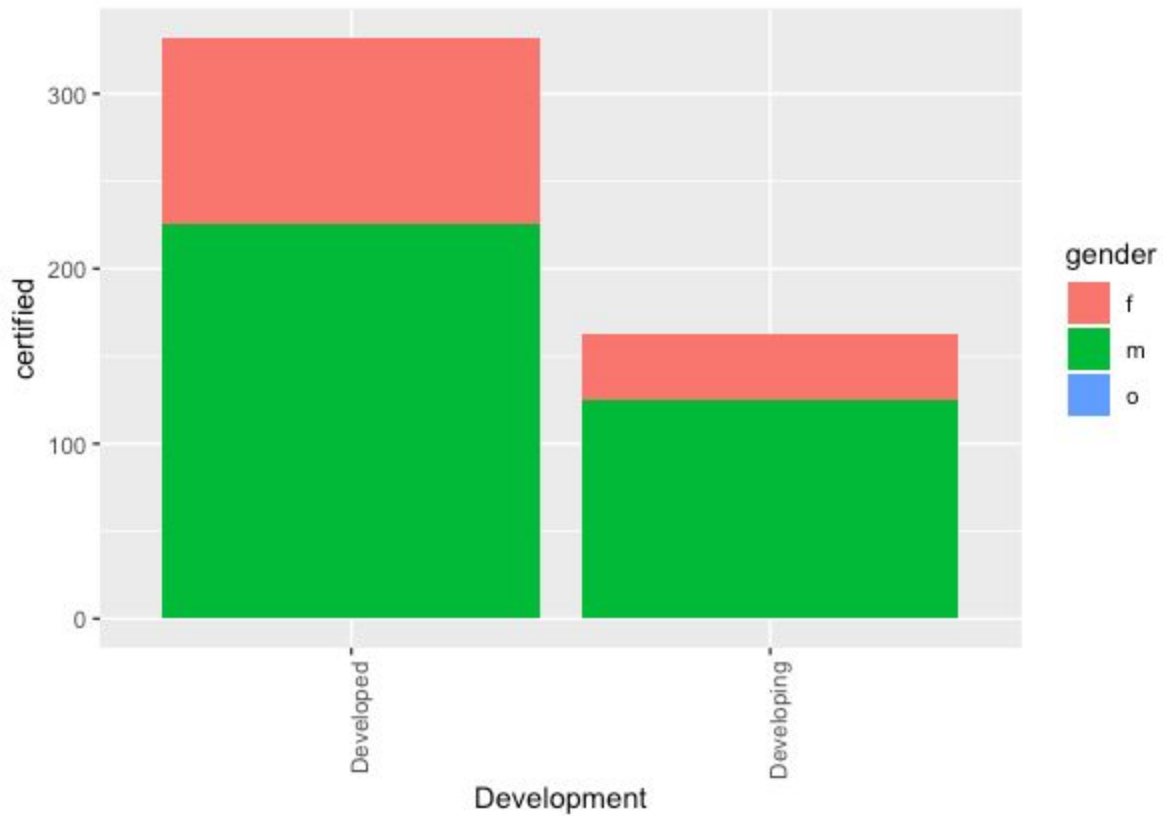
### ***Learning Objective 1: Learner Characteristics***



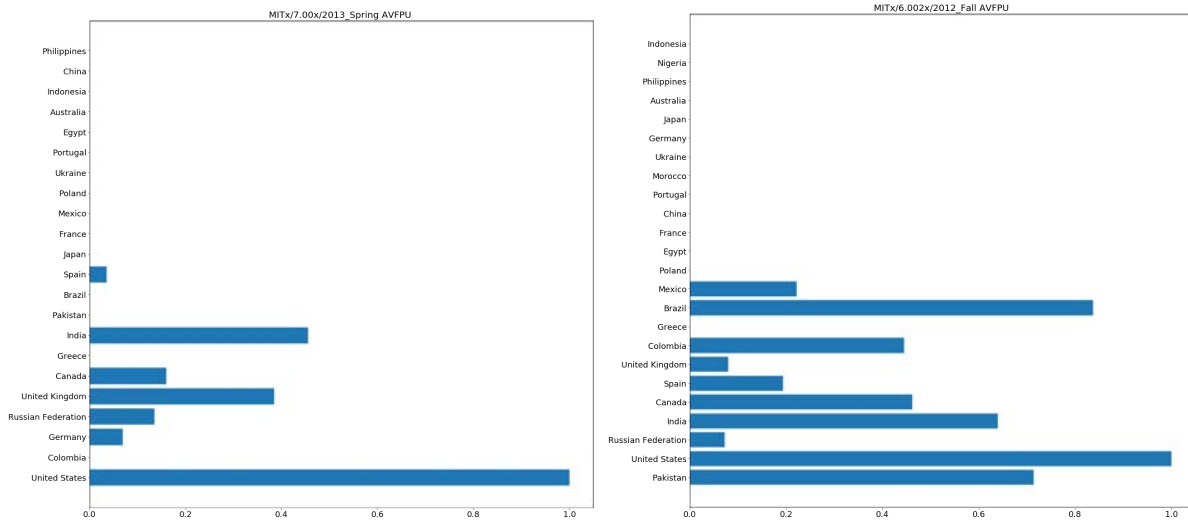
**Figure 1.** Age of certified learners in the edX courses versus country. The countries were further grouped per their development status revealing a clear trend



**Figure 2:** Number of certified users is shown in each country, with female as salmon, male as green, and other as blue.. Learner participation is global but is mostly concentrated in Europe and North America



## Learning Objective 2: Learner Participation



**Figure. 3-4:** AVFPU normalized by dividing by the maximum average. Data from four courses shown for country versus AVFPU. The United States and India often exhibit the highest relative posts.

Users from a number of countries do not post forums at all. A further analysis of whether they use them but do not post would be necessary to understand if they find the forums valuable at all. This study highlights that perhaps course designers should find ways to intervene and encourage the use of forums in certain countries perhaps by encouraging use of local languages or some incentive.

Country	Education Index (2013)	Country	Education Index (2013)
Australia	0.927	Mexico	0.638
Bangladesh	0.447	Morocco	0.468
Brazil	0.661	Nigeria	0.425
Canada	0.850	Pakistan	0.372
China	0.610	Philippines	0.610
Colombia	0.602	Poland	0.825
Egypt	0.573	Portugal	0.728
France	0.816	Russian Federation	0.780
Greece	0.797	Spain	0.794

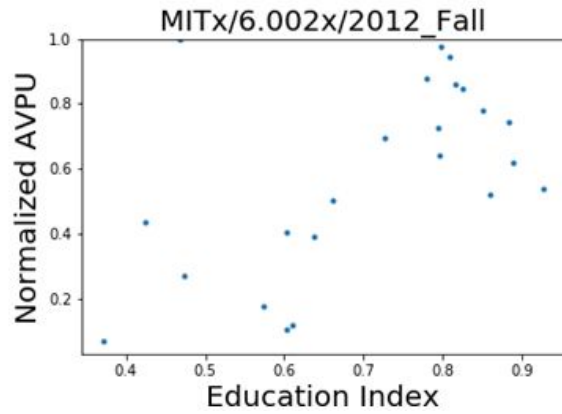
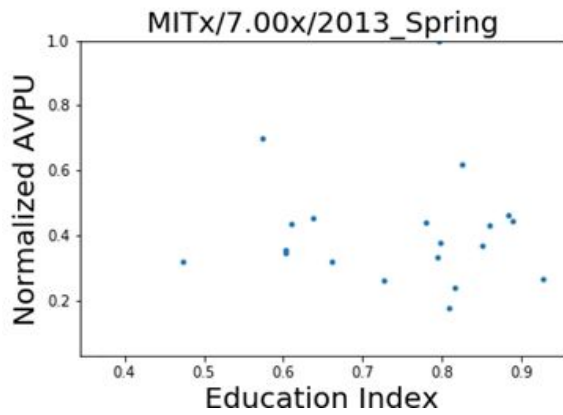
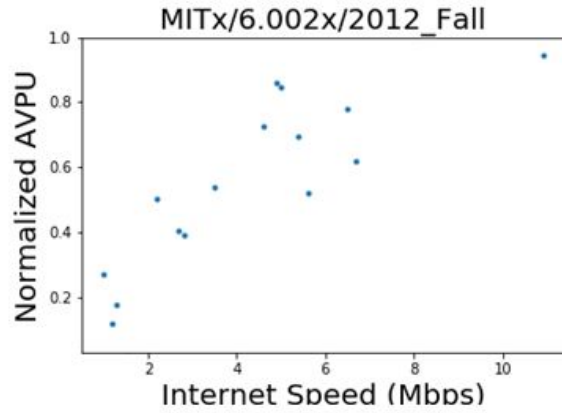
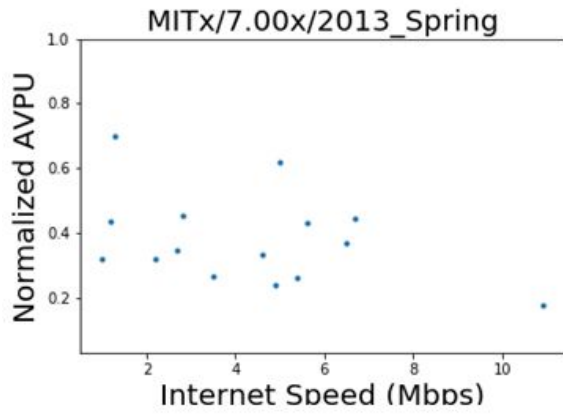
India	0.473	Ukraine	0.796
Indonesia	0.603	United Kingdom	0.860
Japan	0.808	United States	0.890

**Table 1:** Each country listed with its Education Index from 2013. Education Index calculated by UN Human Development Report from the average of the mean years and expected years of schooling. More developed countries tend to have a higher Education Index.

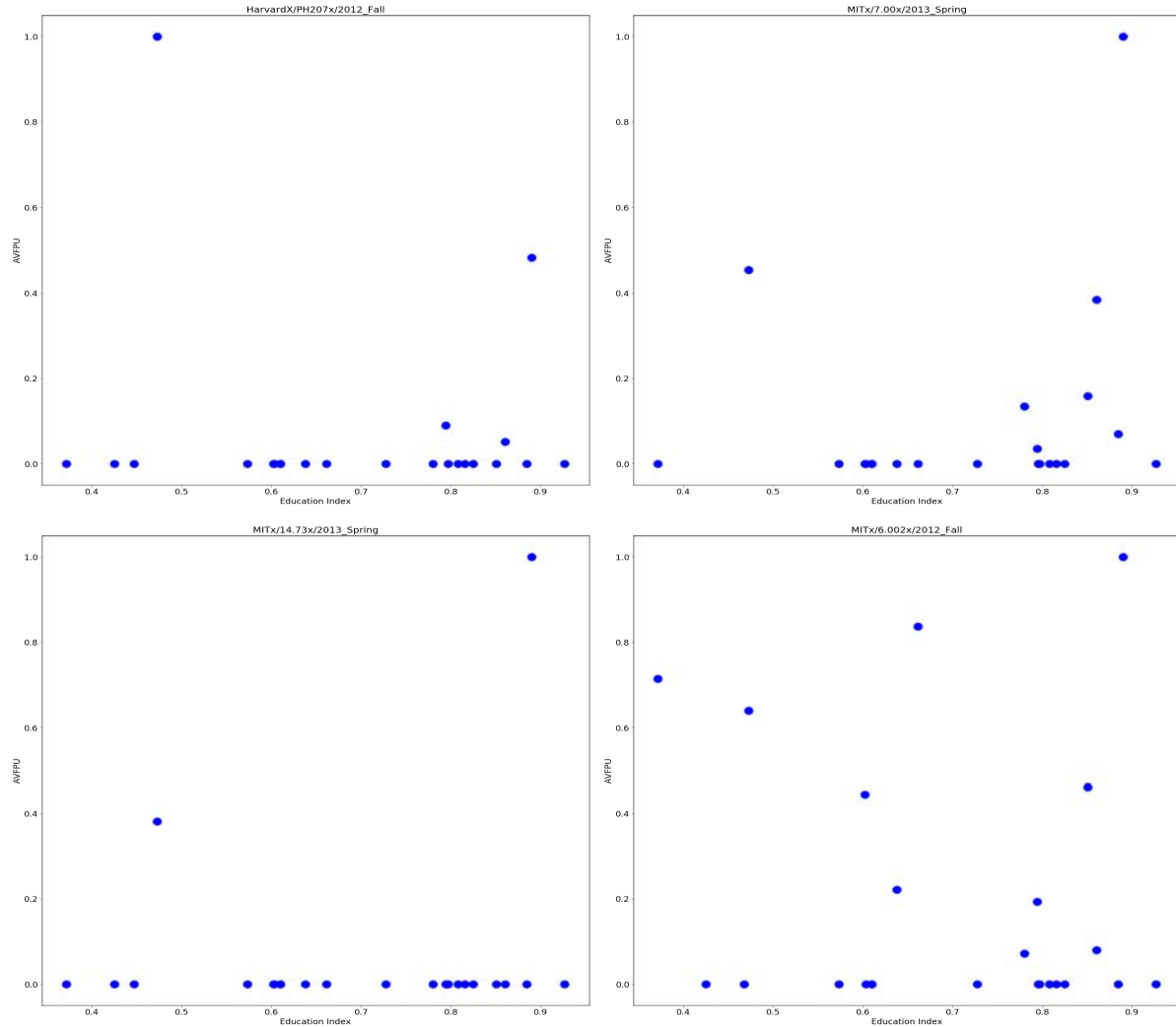
Videos are an integral part of many edX courses and it was found that certified users from all countries interact with them one way or another. The analysis of AVPU against internet speeds produced mixed results with the majority of courses showing no significant trend as shown by the Figure. 5 (top left). Interestingly, both 2012 and 2013 versions of the popular 6.002x course showed visible positive correlations. However, no strong conclusions can be made about whether internet speeds affect video use in other courses so with this analysis, no suggestions can yet be made to course designers. No obvious correlations were found between AVPU and Education Index.

A number of improvements can be made to the analysis including quantifying the correlation using linear regression, including error bars due to comparisons between different sample sizes between countries and using more granular data such as percentage of videos watched as well as the internet speeds for each user as opposed to a national average. Such a future analysis would provide useful insight on whether edX courses are accessible globally.





**Figures 5-8:** Scatterplots of AVPU against Average National Internet Speed in 2012 (Top 2) and Education Index (bottom 2) The results varied from course to course with courses such as 6.002x showing some correlation but most of the other courses showing no obvious correlation as in 7.00x.



**Figure 9-12.** AVFPU versus Education Index for four different courses, showing no correlation in forum activity and Education Index.

## Reproducing your work

The original data for [Education Index](#) and [Internet Speed](#) can be found at the appropriate references listed.

The R script to generate the plots pertaining to course interest can be found [here](#).

```

1. byed<- k%>%
+ ggplot(aes(x= Region, y= certified/sum(registered)*100, fill = LoE_DI))%>% + geom_bar(stat=
"identity") + theme(axis.text.x = element_text(angle = 90, hjust= 1))

2. ggplot(aes(x= EdIndex, y= sum(certified)/sum(registered)*100, color = "Region"))%>% +
geom_point() + theme(axis.text.x = element_text(angle = 90, hjust= 1))

3. genderbydev<- M%>%
+ ggplot(aes(Development, certified, fill = gender ))%>% + geom_bar(stat= "identity") +
theme(axis.text.x = element_text(angle = 90, hjust= 1))

```

## References

- Belson, D. (2010). Akamai state of the internet report, Q1 2012.  
<https://www.akamai.com/us/en/multimedia/documents/state-of-the-internet/akamai-q1-2012-state-of-the-internet-connectivity-report.pdf>. Accessed on 2/26/19
- Alexandron, G., Lee, S., Chen, Z., & Pritchard, D.E. (2016). Detecting Cheaters in MOOCs Using Item Response Theory and Learning Analytics. *UMAP*.
- Human Development Report Office (2013). Education Index.  
<http://hdr.undp.org/en/content/education-index>. Accessed 2/27/19.
- 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).

MIT OpenCourseWare  
<https://ocw.mit.edu>

CMS.594/CMS.894 Education Technology Studio  
Spring 2019

For more information about citing these materials or our Terms of Use, visit:  
<https://ocw.mit.edu/terms>.