



Canvas Network Person-Course (1/2014 - 9/2015) De-Identified Dataset

Created February 29, 2016

Description of the Dataset

This dataset release is comprised of de-identified data from Canvas Network open courses (running January 2014 - September 2015), along with related documentation. These data include over 325,000 aggregate records, and each record represents one individual's activity in one of 238 Canvas Network courses. The structure of this data set is based on [the HarvardX-MITx Person-Course data release of 2014](#) (however, this data release is not affiliated with either Harvard or MIT).

Files:

- CNPC_1401-1509_DI_v1_1_2016-03-01.csv

The md5sum for this release: c769bf2d18f119c7ff7b0fa551a91107

Overview

Canvas Network (<http://canvas.net>) is an open online course website operated by the educational technology company [Instructure](#). Canvas Network hosts open online courses, including Massive Open Online Courses (MOOCs), that are freely available to participants around the world. Data related to the open online courses and the users' experiences in those courses may be beneficial to researchers seeking to better understand and describe the MOOC phenomenon. It may also be useful to gain insight into online learning behaviors and course design characteristics.

Table 1. Course count by discipline in this dataset (1/2014 - 9/2015)

Discipline	Courses
Mathematics & Statistics	7
Professions and Applied Sciences	73
Medical Pre-Medical	5

Education	57
Interdisciplinary & Other	16
Social Sciences	8
Business and Management	28
Humanities	29
Computer Science	9
Physical Sciences	6
SUM	238

This dataset is at the level of one row per-person, per-course. For example, if one individual is enrolled in three Canvas Network courses during the period covered by the dataset, that user would have three rows associated with their user ID. Fields that are course-specific are de-normalized across users; for example, "discipline" and "course_length". This was done to remove the need to join to a course table for analysis.

The dataset includes administrative variables about the Canvas Network courses, variables generated by users during their interaction with the courses, and variables collected through surveys of the users. Many of the variables parallel those published by Harvard and MIT in the 2014 HarvardX-MITx Person-Course data release; we have generalized some variables and added in others. For a detailed description of the variables included in this dataset, please see the variable list under Data Dictionary that follows.

Data Dictionary

Notes

- "Administrative" indicates that the variable comes from the Canvas Network system or has been computed by the research team.
- "User-provided" indicates that the variable comes from questions asked by Canvas Network of the student at the time of account registration or the beginning of the course.
- "H-M" indicates that this variable matches a variable as defined by the HarvardX-MITx data release of 2014 as closely as possible.
- Blank values should be read as null. These typically indicate that no data was available for this variable, either because the user did not interact with the system in such a way as to produce data, or because the course itself was not designed in a such a way to produce this data.
- "_DI" at the end of a variable name indicates that this variable was transformed during the de-identification process.

- "Withheld" indicates a variable that needed to be dropped in the current version in order to ensure de-identification, but may be added in a future release after additional tooling.

Columns

course_id_DI : administrative, string, a unique identifier for the course, randomly generated
discipline: administrative, string, a generalization of the course title as the discipline of the course, derived from Canvas Network's internal classification.

userid_DI : H-M, administrative, string, a unique identifier for the user in this dataset, randomly generated

registered : H-M, administrative, 0/1, registered for course, =1 for all records in person-course.

viewed : H-M, administrative, 0/1, =1 if the number of interactions within the course are greater than 1.

explored : H-M, administrative, 0/1, =1 if the user interacted with or viewed $\geq 50\%$ of the course modules (i.e. chapters).

completed_% : administrative, 0.0 - 1.0, percent of total required content modules completed if number of content modules with requirements is greater than an internal threshold (else blank).

course_reqs : administrative, 0/1, =1 if ≥ 3 content modules with requirements
grade: H-M, administration, 0.0 - 1.0, estimated or actual final grade in the course as a percent, for courses that have sufficient scored activities to calculate a final grade.

grade_reqs : administrative, 0/1 =1 if ≥ 3 assignments in course

primary_reason : user-provided, string, standardized reason for taking course (from course survey)

final_cc_cname_DI : H-M, user-provided, withheld

primary_reason : user-provided, string, standardized reason for taking course (from course survey)

learner_type : user-provided, string, standardized type of learner (from course survey)

expected_hours_week : user-provided, string, standardized range of hours per week (from course survey)

LoE : user-provided, highest level of education completed. Possible values: Level of Education (from survey)

age_DI : user-provided. Age brackets.

gender : H-M, user-provided, withheld. Possible values: m (male), f (female), other, blank (not-specified)

start_time_DI : administrative, quarter and year that the first user interaction occurred (binned from actual date).

course_start : administrative, quarter and year that the course officially started (binned from actual date).

course_end : administrative, quarter and year that the course officially ended (binned from actual date).

last_event_DI : administrative, quarter and year that the last user interaction occurred (binned from actual date).

nevents : H-M, administrative, count of distinct interactions with the course, recorded as page views.

ndays_act : H-M, administrative, count of distinct days with one or more *events*.

nchapters : administrative, 0.0 - 1.0, percent of unique modules (i.e. chapters) viewed in course

nforum_posts : H-M, administrative, number of posts total in discussion forums throughout the course.

course_length : administrative, number of days that course officially ran or that course had participant activity, whichever is larger (binned).