Hatch, Match and Dispatch:

Examining the relationship between student intent, expectations, behaviours and outcomes in six Coursera MOOCs at the University of Toronto

UNIVERSITY OF TORONTO

The Team

Bodong Chen (@bodongchen)

Stian Håklev (@houshuang)

William Heikoop

Laurie Harrison (@IT4learning)

Hedieh Najafi

Carol Rolheiser

Chris Teplovs (@cteplovs)

Research Objectives

In the context of 6 University of Toronto MOOCs, we examined:

Student intentions, goals and motivation (Hatch) and their relationship (Match) to student behaviour and student performance based on formative and summative assessment (Dispatch)

Research Design

Multiple case study of six MOOCs:

- Aboriginal Worldviews and Education
- Introduction to Psychology
- Learn to Program: The Fundamentals
- Learn to Program: Crafting Quality Code
- The Social Context of Mental Health and Illness
- Statistics: Making Sense of Data

Research Design

Multiple case study of six MOOCs:

- Aboriginal Worldviews and Education
- **Introduction to Psychology**
- Learn to Program: The Fundamentals
- Learn to Program: Crafting Quality Code
- The Social Context of Mental Health and Illness
- Statistics: Making Sense of Data

Data Sources

1. Entry Survey (demographic info, motivation, expected engagement)
2. Student behaviour data (clickstream data)
3. Outcome data (quiz scores, final grades)
4. Exit survey
5. Instructor interviews

Data Sources

1. **Entry Survey** (demographic info, **motivation, expected engagement**)
2. **Student behavior data (clickstream data)**
3. Outcome data (quiz scores, **final grades**)
4. Exit survey
5. Instructor interviews

Data Analysis

Part I: Entry survey (n=70,418 for all MOOCs)

- Principal component analysis (PCA) of surveys to reduce complexity of categorization.
- Two axes: “betterment” and “enjoyment”

Part II: Clickstream data (n=16,900,926 events for Introductory Psychology only)

- Sequential pattern mining

Part III: Comparing frequent patterns from Part II based on classification from Part I

Data Analysis Part Ia: Entry Surveys

This survey is intended for learners using this course as an "archived" MOOC. We consider an "archived" MOOC to be a course in which course materials are available, but for which deadlines have passed. The alternative is a "live" MOOC for which the due dates are upcoming. Your feedback will help us understand how our MOOCs can be better designed to meet the diverse needs and interests of our learners.

1. Which of the following descriptions best characterizes you?

- ☐ Student
- ☐ Professional
- ☐ Researcher
- ☐ Academic / Professor
- ☐ Computer Engineering
- ☐ Lifelong Learner
- ☐ None of the above

2. Why did you enroll in this course? For each reason below, please rate on the scale of not at all important to very important.

	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
The subject is relevant to my academic field of study					
This class teaches skills that will help my job/career					
I want to earn some sort of credential that I can use to enhance my CV / resume					
Because this course is offered by a prestigious university					
I think taking this course will be fun and enjoyable					

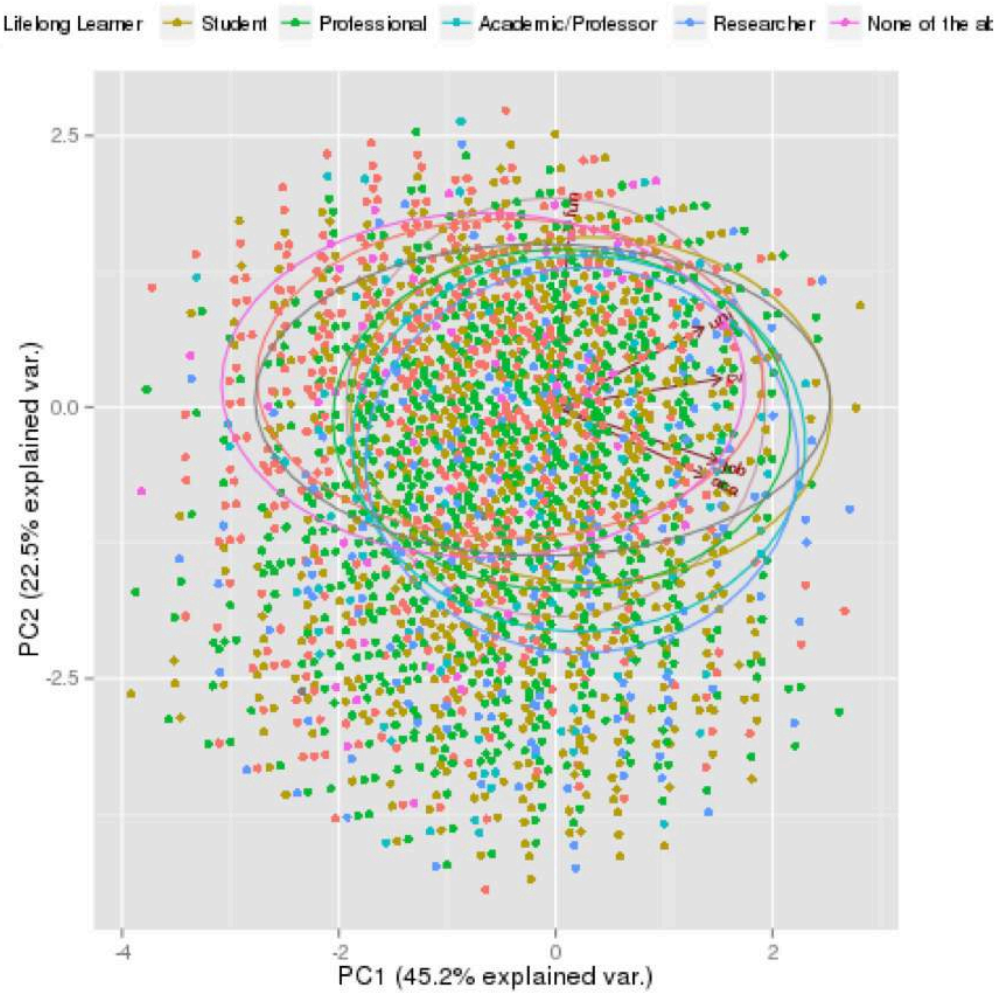
Data Analysis Part Ib: Principal Components Analysis of Survey Data

```
##         PC1     PC2     PC3      PC4      PC5
## aca  0.4782 -0.3096  0.4917 -0.53174 -0.38852
## job  0.5187 -0.2458  0.3214  0.48693  0.57459
## cv   0.5273  0.1293 -0.4000  0.45436 -0.58206
## uni  0.4714  0.3607 -0.4558 -0.51393  0.41923
## fun  0.0447  0.8348  0.5359  0.09786 -0.06596
```

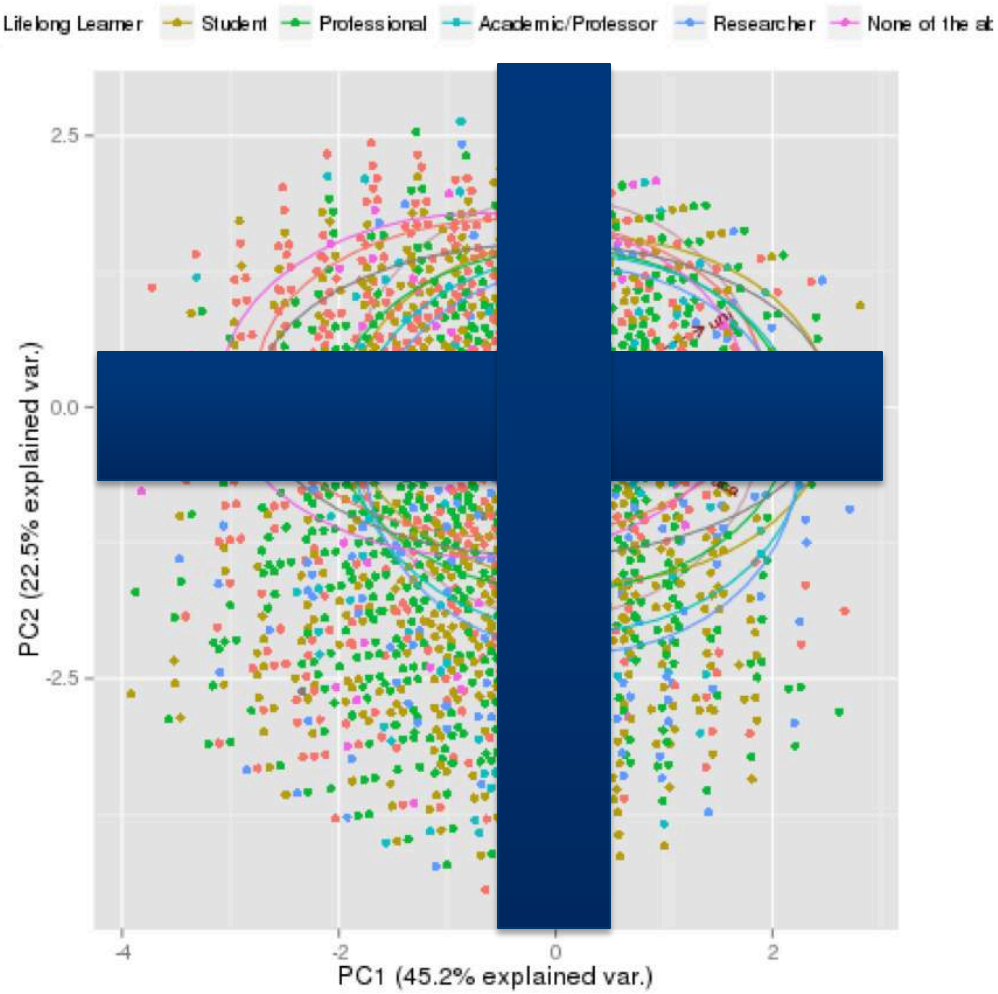
2. Why did you enroll in this course? For each reason below, please rate on the scale of not at all important to very important.

	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
The subject is relevant to my academic field of study					
This class teaches skills that will help my job/career					
I want to earn some sort of credential that I can use to enhance my CV / resume					
Because this course is offered by a prestigious university					
I think taking this course will be fun and enjoyable					

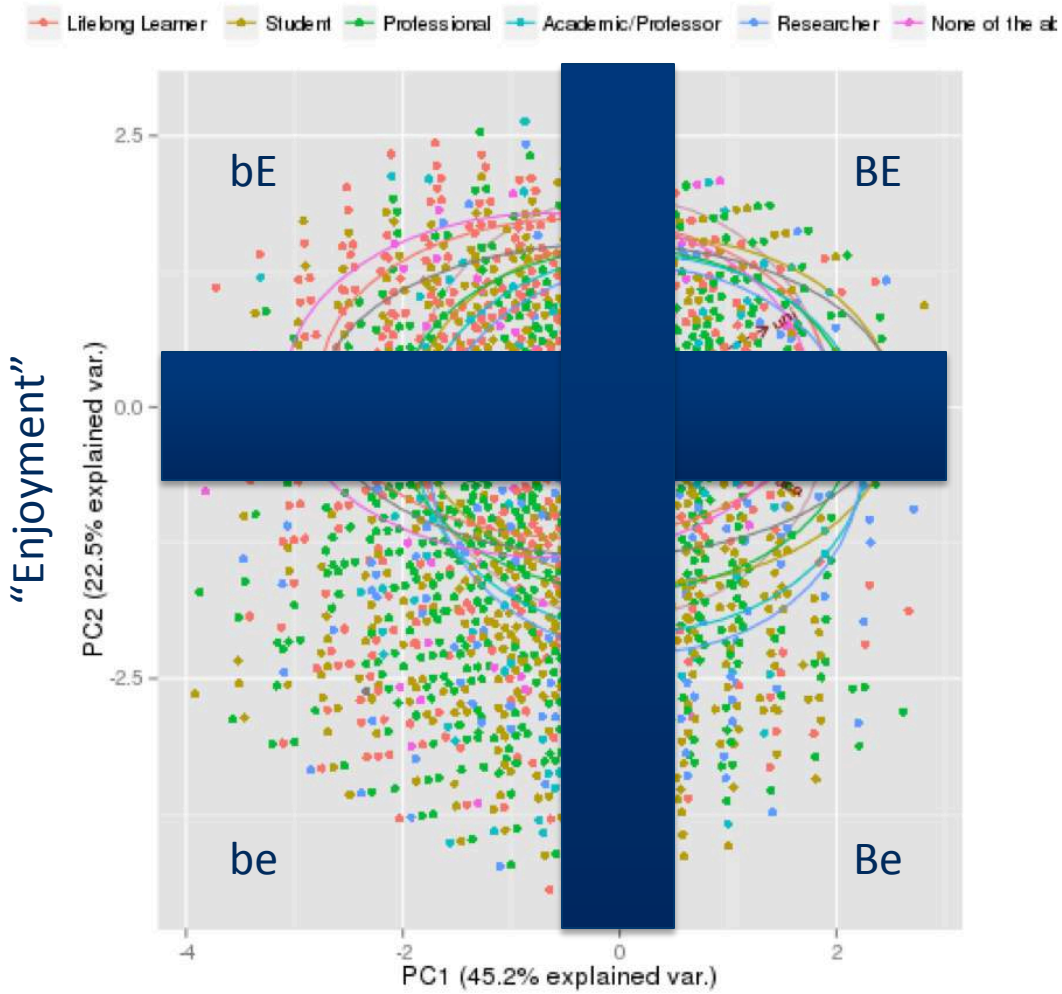
Data Analysis Part Ic: PCA continued



Data Analysis Part Ic: PCA continued



Data Analysis Part Ic: PCA continued



Data Analysis Part IIa: Clickstream Data

3 Clickstream data export

The clickstream data are given as a single gzipped text file. The text file contains a listing of clickstream events, one per line. Each entry consists of a single JSON-serialized object with the following fields:

- key: a string describing a particular kind of event
- value: a string containing metadata for the event
- username: anonymized user ID (`anon_user_id`; missing if individual not logged in)
- timestamp: POSIX timestamp indicating when event occurred
- page_url: the webpage associated with an event
- client: a keyword describing the context of an event
- session: browser session cookie
- language: the client browser's language preference
- from: the referer URL
- user_ip: IP address of user
- user_agent: browser user agent string

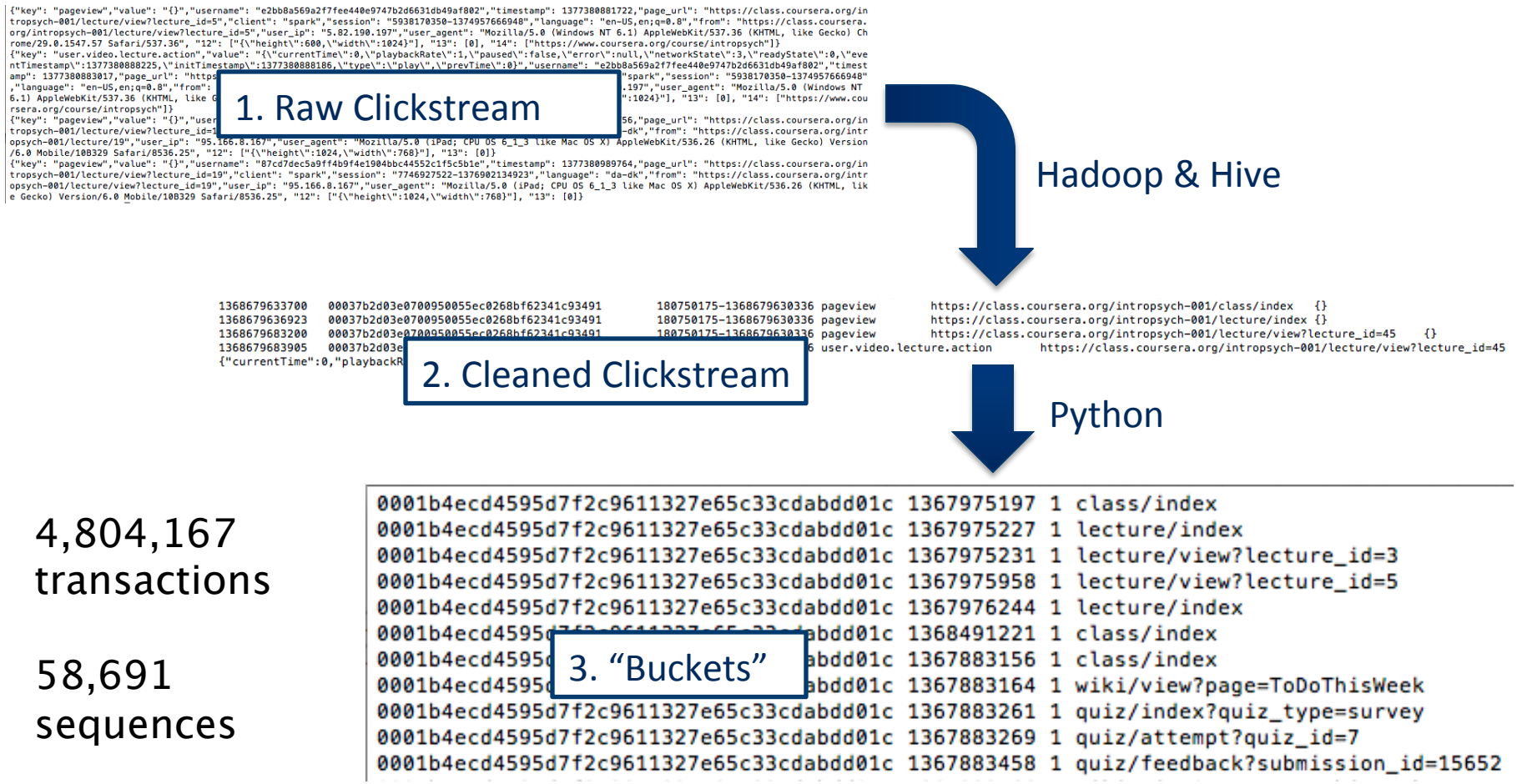
In general, the export contains two types of events: video-related events and page view events. In both cases, the value field itself is a JSON-serialized object that contains metadata associated with the event. For example, for video events, the value field contains information about the specific type of action taken (e.g., play, pause, seek) and current video settings (e.g., current playback speed).

Info from Coursera

This is what clickstream data looks like


```
{"key": "pageview","value": "{}","username": "e2bb8a569a2f7fee440e9747b2d6631db49af802","timestamp": 1377380881722,"page_url": "https://class.coursera.org/intropsych-001/lecture/view?lecture_id=5","client": "spark","session": "5938170350-1374957666948","language": "en-US,en;q=0.8","from": "https://class.coursera.org/intropsych-001/lecture/view?lecture_id=5","user_ip": "5.82.190.197","user_agent": "Mozilla/5.0 (Windows NT 6.1) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/29.0.1547.57 Safari/537.36", "12": ["{\"height\":600,\"width\":1024}"], "13": [0], "14": ["https://www.coursera.org/course/intropsych"]}
{"key": "user.video.lecture.action","value": "{\"currentTime\":0,\"playbackRate\":1,\"paused\":false,\"error\":null,\"networkState\":3,\"readyState\":0,\"eventTimestamp\":1377380888225,\"initTimestamp\":1377380888186,\"type\":\"play\",\"prevTime\":0}","username": "e2bb8a569a2f7fee440e9747b2d6631db49af802","timestamp": 1377380883017,"page_url": "https://class.coursera.org/intropsych-001/lecture/view?lecture_id=5","client": "spark","session": "5938170350-1374957666948","language": "en-US,en;q=0.8","from": "https://class.coursera.org/intropsych-001/lecture/5","user_ip": "5.82.190.197","user_agent": "Mozilla/5.0 (Windows NT 6.1) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/29.0.1547.57 Safari/537.36", "12": ["{\"height\":600,\"width\":1024}"], "13": [0], "14": ["https://www.coursera.org/course/intropsych"]}
{"key": "pageview","value": "{}","username": "87cd7dec5a9ff4b9f4e1904bbc44552c1f5c5b1e","timestamp": 1377380989756,"page_url": "https://class.coursera.org/intropsych-001/lecture/view?lecture_id=19","client": "spark","session": "7746927522-1376902134923","language": "da-dk","from": "https://class.coursera.org/intropsych-001/lecture/19","user_ip": "95.166.8.167","user_agent": "Mozilla/5.0 (iPad; CPU OS 6_1_3 like Mac OS X) AppleWebKit/536.26 (KHTML, like Gecko) Version/6.0 Mobile/10B329 Safari/8536.25", "12": ["{\"height\":1024,\"width\":768}"], "13": [0]}
{"key": "pageview","value": "{}","username": "87cd7dec5a9ff4b9f4e1904bbc44552c1f5c5b1e","timestamp": 1377380989764,"page_url": "https://class.coursera.org/intropsych-001/lecture/view?lecture_id=19","client": "spark","session": "7746927522-1376902134923","language": "da-dk","from": "https://class.coursera.org/intropsych-001/lecture/view?lecture_id=19","user_ip": "95.166.8.167","user_agent": "Mozilla/5.0 (iPad; CPU OS 6_1_3 like Mac OS X) AppleWebKit/536.26 (KHTML, like Gecko) Version/6.0 Mobile/10B329 Safari/8536.25", "12": ["{\"height\":1024,\"width\":768}"], "13": [0]}
```

Data Analysis Part IIb: Clickstream Data Wrangling Workflow



Data Analysis Part IIc: Clickstream Analysis (Sequential Pattern Mining)

```
Console ~/

Restarting R session...

Loading required package: arulesSequences
Loading required package: arules
Loading required package: Matrix
Loading required package: lattice

Attaching package: 'arules'

The following objects are masked from 'package:base':

    %in%, write

> load("~/.RData")
> summary(ip1p1fs)
set of 60 sequences with

most frequent items:
     quiz/attempt?quiz_id=7 wiki/view?
                         16
  lecture/view?lecture_id=3
                          8

most frequent elements:
     {quiz/attempt?quiz_id=7} {wiki/vi
                           16
              {lecture/index}   {lectu
                            8

element (sequence) size distribution:
sizes
 1  2  3
24 35  1

sequence length distribution:
lengths
 1  2  3
24 35  1

summary of quality measures:
    support
 Min.   :0.5012
 1st Qu.:0.5398
 Median :0.5875
 Mean   :0.6287
 3rd Qu.:0.6594
 Max.   :0.9951

mining info:
  data ntransactions nsequences suppor
 ip1p1        293984       1622     0.
>
```

Using the 'arulesSequences' package in R, which implements Zaki (2001)...

M. J. Zaki. (2001). SPADE: An Efficient Algorithm for Mining Frequent Sequences. Machine Learning Journal, 42, 31–60.

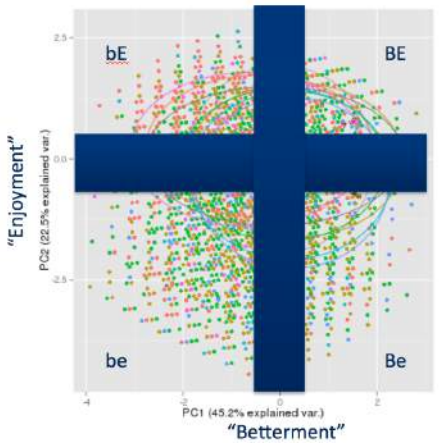
```
Console ~/
> as(ip1p1fs,"data.frame")
                                                         sequence   support
1                                                 <{class/index}> 0.9950678
2                                                 <{forum/index}> 0.6874229
3                                        <{forum/list?forum_id=2}> 0.7774353
4                                                <{lecture/39}> 0.5425401
5                                               <{lecture/index}> 0.9543773
6                                    <{lecture/view?lecture_id=3}> 0.7953144
7                                   <{lecture/view?lecture_id=37}> 0.5012330
8                                   <{lecture/view?lecture_id=39}> 0.6337855
9                                   <{lecture/view?lecture_id=47}> 0.5400740
10                                   <{lecture/view?lecture_id=5}> 0.6972873
11                                   <{lecture/view?lecture_id=7}> 0.5838471
12                                   <{lecture/view?lecture_id=9}> 0.5246609
13                                  <{quiz/attempt?quiz_id=49}> 0.5505549
14                                  <{quiz/attempt?quiz_id=51}> 0.6602959
15                                  <{quiz/attempt?quiz_id=55}> 0.6732429
16                                   <{quiz/attempt?quiz_id=7}> 0.9938348
17                             <{quiz/index?quiz_type=survey}> 0.8378545
18                                    <{quiz/start?quiz_id=49}> 0.5856967
19                     <{wiki/view?page=AnsweringQuestions}> 0.6171393
20                        <{wiki/view?page=courselogistics}> 0.5110974
21                          <{wiki/view?page=DigitalLabCoat}> 0.7435265
22                                   <{wiki/view?page=mTuner}> 0.6288533
23                                 <{wiki/view?page=pScholar}> 0.5579531
24                               <{wiki/view?page=ToDoThisWeek}> 0.9654747
25                 <{class/index},{wiki/view?page=ToDoThisWeek}> 0.5086313
26       <{forum/list?forum_id=2},{wiki/view?page=ToDoThisWeek}> 0.6183724
27    <{lecture/view?lecture_id=3},{wiki/view?page=ToDoThisWeek}> 0.5955610
28      <{quiz/attempt?quiz_id=7},{wiki/view?page=ToDoThisWeek}> 0.8249075
29  <{quiz/index?quiz_type=survey},{wiki/view?page=ToDoThisWeek}> 0.6048089
30   <{wiki/view?page=ToDoThisWeek},{wiki/view?page=ToDoThisWeek}> 0.5209618
31            <{quiz/attempt?quiz_id=7},{wiki/view?page=mTuner}> 0.5425401
32     <{forum/list?forum_id=2},{wiki/view?page=DigitalLabCoat}> 0.5104809
33  <{lecture/view?lecture_id=3},{wiki/view?page=DigitalLabCoat}> 0.5030826
34    <{quiz/attempt?quiz_id=7},{wiki/view?page=DigitalLabCoat}> 0.6578298
35 <{quiz/attempt?quiz_id=7},{wiki/view?page=AnsweringQuestions}> 0.5314427
```

...we extract frequently occurring sequences

Early Results

We compare different categories, similar to approach taken by Sabourin, Mott & Lester, 2013

Sequence	Partition				Spread
	BE	Be	be	bE	
<{class/index},{class/index}>	51%	53%	54%	52%	3%
<{class/index},{lecture/index}>	54%	55%	57%	57%	3%
<{class/index},{wiki/view?page=ToDoThisWeek}>	51%		53%	55%	4%
<{forum/list?forum_id=2},{class/index}>	60%	52%	51%	53%	9%
<{forum/list?forum_id=2},{lecture/index}>	66%	57%	56%	55%	11%
<{forum/list?forum_id=2},{quiz/index?quiz_type=survey}>	54%				0%
<{forum/list?forum_id=2},{wiki/view?page=DigitalLabCoat}>	51%				0%
<{forum/list?forum_id=2},{wiki/view?page=ToDoThisWeek}>	62%	55%	52%		10%
<{lecture/index},{class/index}>			51%		0%
<{lecture/index},{lecture/index}>			52%		0%
<{lecture/view?lecture_id=3},{class/index}>	59%	56%	59%		4%
<{lecture/view?lecture_id=3},{lecture/index}>	60%	58%	61%		3%
d=3},{lecture/view?lecture_id=5},{lecture/view?lecture_id=7}>			54%		0%
<{lecture/view?lecture_id=3},{lecture/view?lecture_id=5}>	62%	61%	65%	59%	6%
<{lecture/view?lecture_id=3},{lecture/view?lecture_id=7}>	53%	53%	57%	52%	5%
<{lecture/view?lecture_id=3},{lecture/view?lecture_id=9}>			51%		0%
{lecture/view?lecture_id=3},{wiki/view?page=DigitalLabCoat}>	50%		50%		0%
lecture/view?lecture_id=3},{wiki/view?page=ToDoThisWeek}>	60%	56%	58%	52%	7%
<{lecture/view?lecture_id=5},{class/index}>			52%		0%
<{lecture/view?lecture_id=5},{lecture/index}>			53%		0%
<{lecture/view?lecture_id=5},{lecture/view?lecture_id=7}>	54%	51%	57%	52%	6%
<{lecture/view?lecture_id=5},{lecture/view?lecture_id=9}>			51%		0%
lecture/view?lecture_id=5},{wiki/view?page=ToDoThisWeek}>			50%		0%
<{lecture/view?lecture_id=7},{lecture/view?lecture_id=9}>			51%		0%
<{quiz/attempt?quiz_id=51},{lecture/index}>			50%		0%
<{quiz/attempt?quiz_id=55},{lecture/index}>			50%		0%
<{quiz/attempt?quiz_id=7},{class/index}>	79%	78%	80%	75%	5%
<{quiz/attempt?quiz_id=7},{forum/index}>	58%	61%	61%	54%	8%
<{quiz/attempt?quiz_id=7},{lecture/index}>	83%	84%	86%	83%	4%
<{quiz/attempt?quiz_id=7},{lecture/view?lecture_id=3}>	58%	60%	61%	55%	6%
<{quiz/attempt?quiz_id=7},{lecture/view?lecture_id=39}>	55%	53%	56%		3%
<{quiz/attempt?quiz_id=7},{lecture/view?lecture_id=5}>	55%	57%	58%	52%	6%
<{quiz/attempt?quiz_id=7},{lecture/view?lecture_id=7}>			52%		0%
<{quiz/attempt?quiz_id=7},{quiz/attempt?quiz_id=49}>			52%		0%
<{quiz/attempt?quiz_id=7},{quiz/attempt?quiz_id=51}>	60%	61%	62%	58%	4%
<{quiz/attempt?quiz_id=7},{quiz/attempt?quiz_id=55}>	61%	61%	63%	57%	6%
<{quiz/attempt?quiz_id=7},{quiz/index?quiz_type=survey}>	57%	60%	61%	58%	3%
<{quiz/attempt?quiz_id=7},{quiz/start?quiz_id=49}>	52%		53%		2%
<{quiz/attempt?quiz_id=7},{wiki/view?page=AnsweringQuestions}>	53%	52%	55%		3%
<{quiz/attempt?quiz_id=7},{wiki/view?page=DigitalLabCoat}>	66%	63%	65%	59%	7%
<{quiz/attempt?quiz_id=7},{wiki/view?page=mTuner}>	54%	54%	55%		2%
<{quiz/attempt?quiz_id=7},{wiki/view?page=ToDoThisWeek},{lecture/index}>	51%		54%	50%	4%
<{quiz/attempt?quiz_id=7},{wiki/view?page=ToDoThisWeek}>	82%	83%	84%	79%	5%
<{quiz/index?quiz_type=survey},{class/index}>	58%	60%	60%	54%	6%
<{quiz/index?quiz_type=survey},{lecture/index}>	62%	64%	65%	61%	4%
<{quiz/index?quiz_type=survey},{quiz/attempt?quiz_id=7}>				52%	0%
<{quiz/index?quiz_type=survey},{wiki/view?page=ToDoThisWeek}>	60%	59%	61%	57%	4%
<{quiz/start?quiz_id=49},{quiz/attempt?quiz_id=49}>			52%		0%
<{wiki/view?page=ToDoThisWeek},{class/index}>	54%	55%	57%	53%	4%
<{wiki/view?page=ToDoThisWeek},{lecture/index}>	59%	57%	62%	60%	5%
<{wiki/view?page=ToDoThisWeek},{wiki/view?page=ToDoThisWeek}>	52%	51%	54%	53%	3%



Comparing the prevalence of frequently occurring sequences across the four partitions (BE, Be, bE, and be), we see that "be" shows more "return to outline" behaviours and less engagement with forums.

Number of Cases in Each Partition

Number of Cases in Each Partition

Outcome	Intent		Aboriginal Education	Intro Psych	Intro Stats	Mental Health 1	Mental Health 2	Programming 2
	Betterment	Enjoyment						
Certificate	High	High	283	332	227	290	182	634
No Certificate	High	High	27	1400	1458	54	881	861
Certificate	High	Low	234	234	642	317	206	624
No Certificate	High	Low	25	951	3821	38	1058	1221
Certificate	Low	High	328	973	155	162	120	353
No Certificate	Low	High	36	3702	1008	47	990	441
Certificate	Low	Low	215	533	642	106	121	438
No Certificate	Low	Low	49	2350	4777	44	906	736

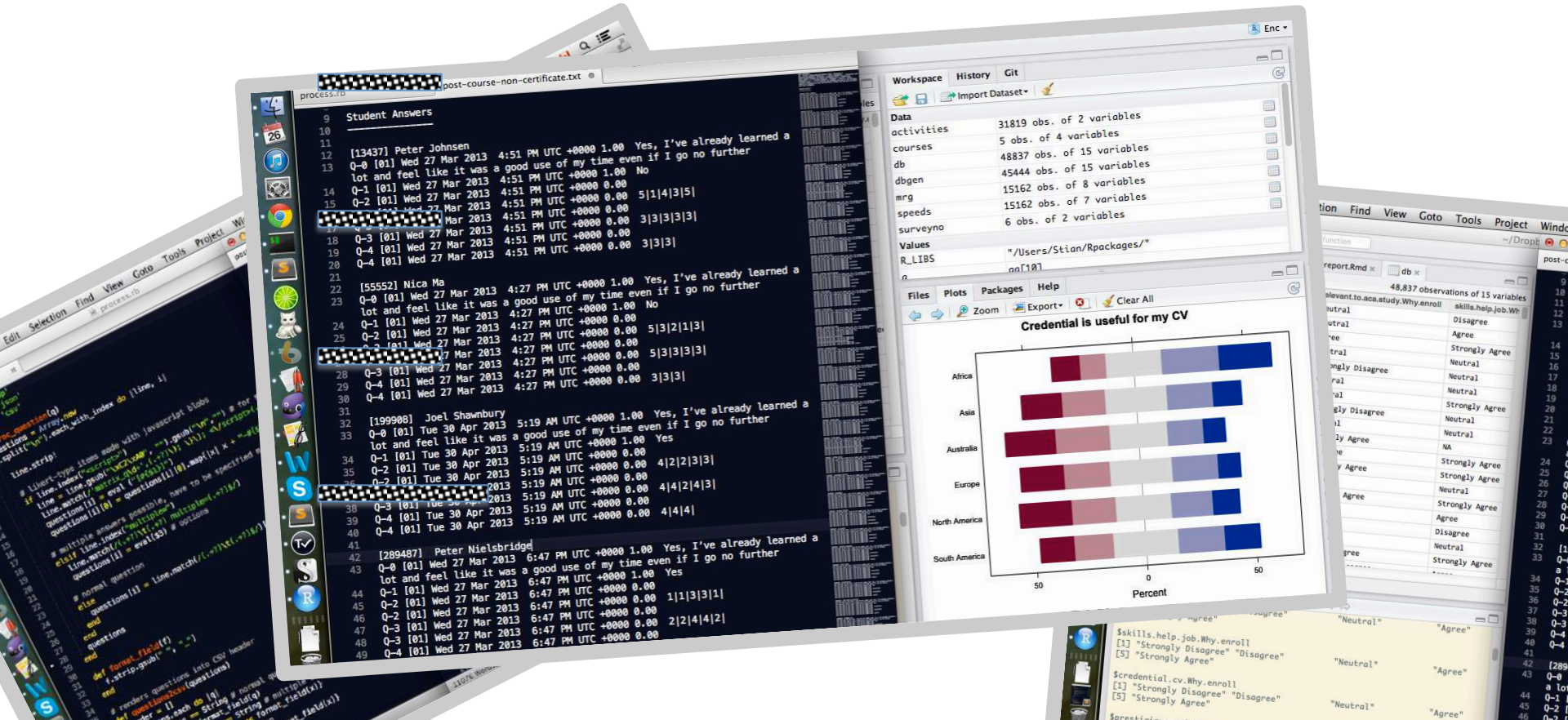
Most Frequent Sequences

- Certificate
  - Common: {wiki/view}{lecture/view}{short-event}{class/index}
  - Exclusive to Low Betterment, Low Enjoyment: {lecture/view, out-of-sequence}{lecture/index}
- No-Certificate
  - Common: {class/index}{lecture/index}{quiz/attempt}{lecture/view}
  - Exclusive to High Betterment, High Enjoyment: {wiki/view}

Data Wrangling and Analysis: A Look Behind the Scenes

Stian Haklev

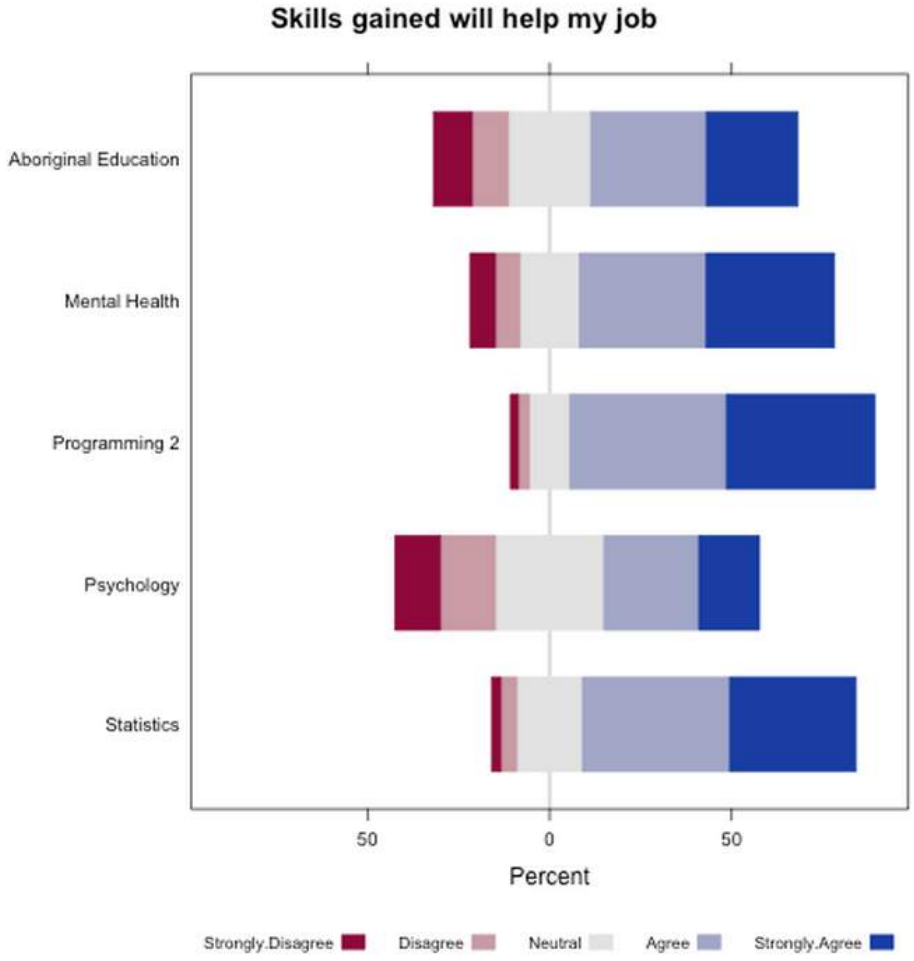
Open UToronto Institutional Researcher

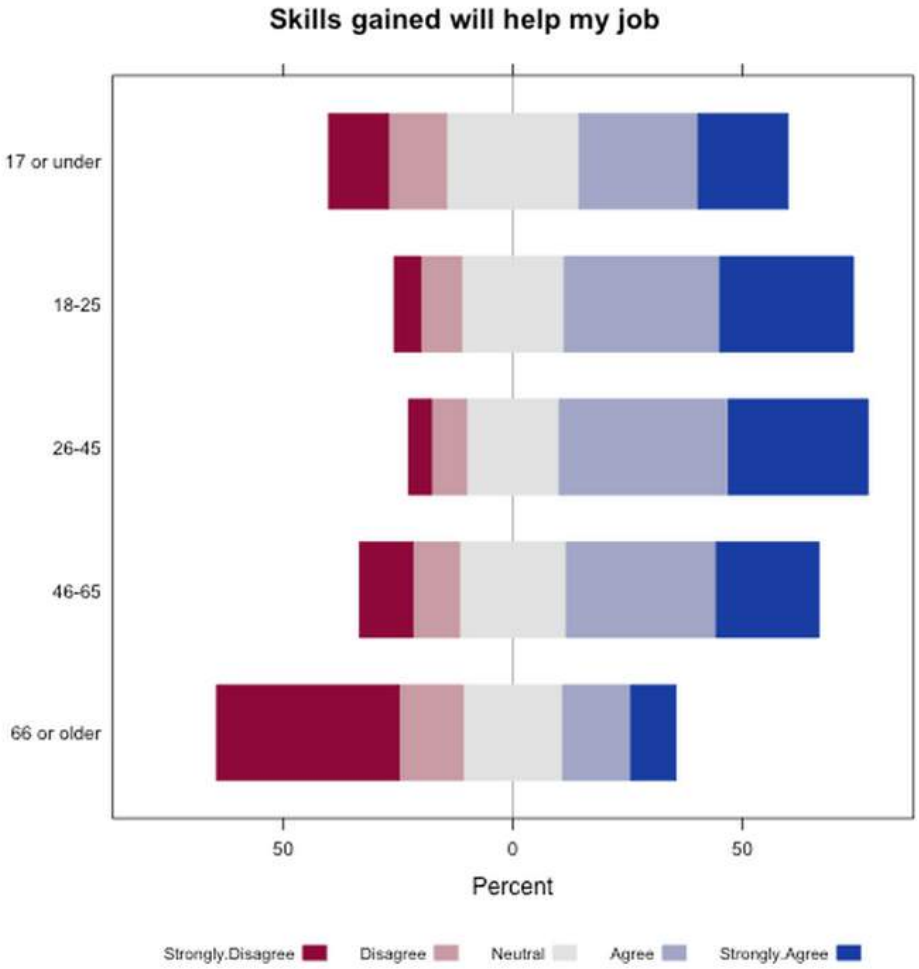


Challenges & What We Are Learning

1. Preparing and combining complex data structures for analysis
2. Managing secure and anonymised access to multiple research teams/researchers/ethics agreements
3. Securing enough computing capacity for “big data” analysis
4. Conducting collaborative research in a documented, reproducible way

Simple relationships from survey data





Demographic Reports

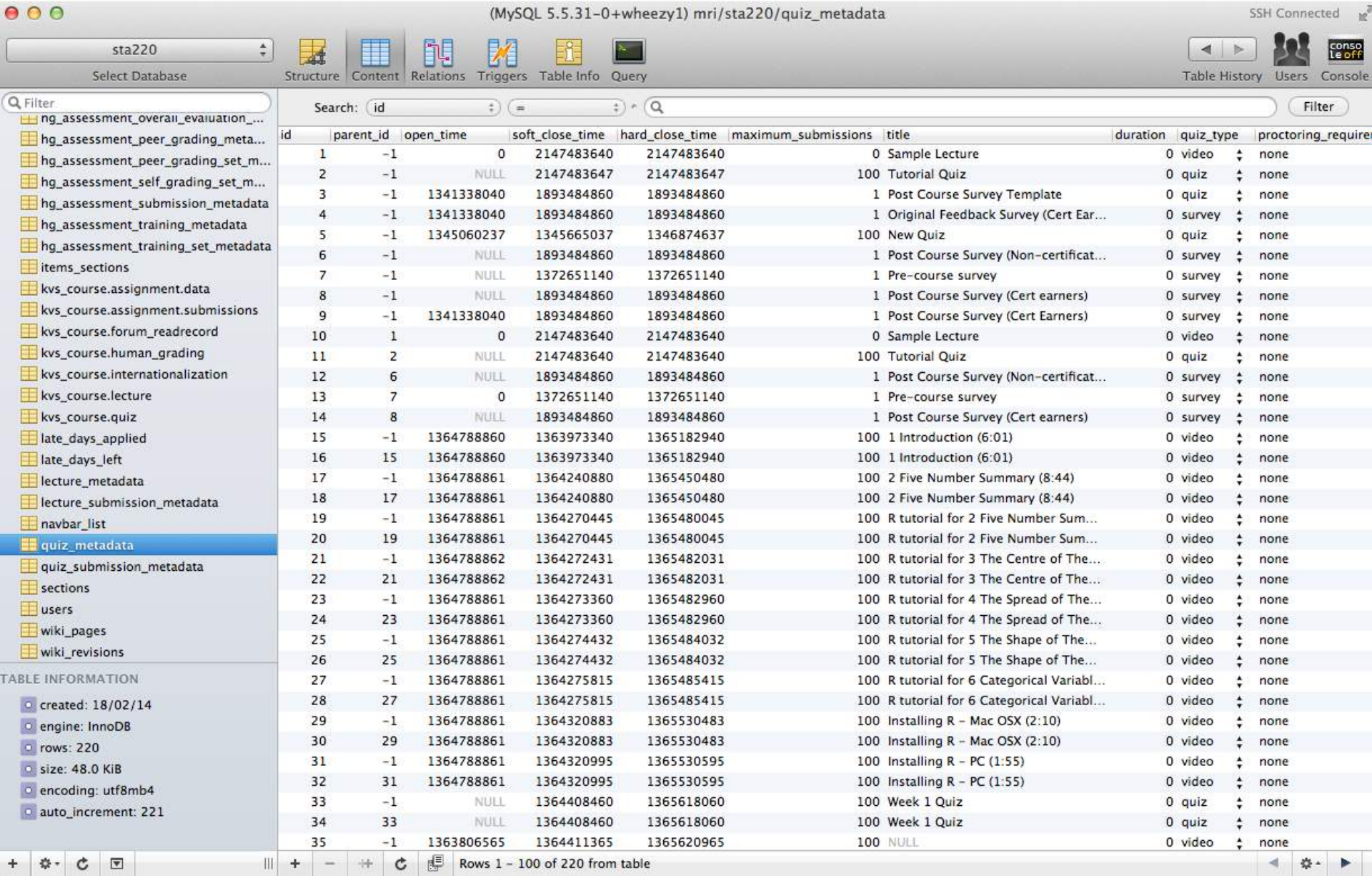
Demographic Report on University of Toronto Coursera MOOCs

June 16, 2013

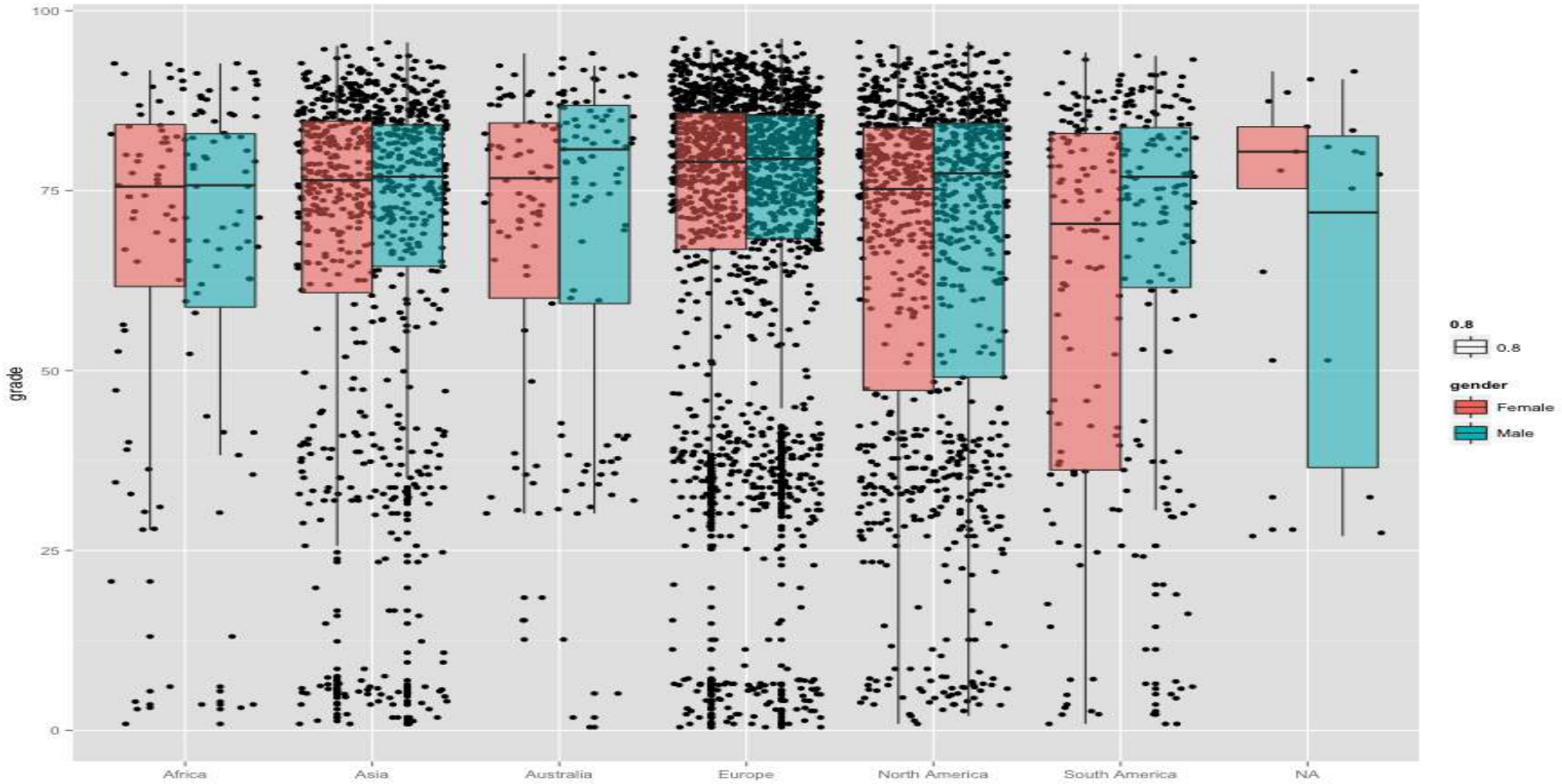
Background

In order to support MOOC-related research and evaluation processes, the Office of Online Learning Strategies is involved in administrating access to detailed user data sets from Coursera and cleaning it for analysis. The data is being stored centrally and made available for a range of purposes, including institutional planning processes. It will also be used in a research study among five MOOCs that launched between January and May 2013. The course instructors are all co-PIs on the ethics proposal, and will share the analytics data. For a full description of activities visit: Open Utoronto MOOC Research and Evaluation.

Connecting database with other data



Integrated surveys allow us to link intention and demographics with outcomes and behaviour



Tools to collaborate and document

IP[y]: Notebook Untitled0 Last Checkpoint: Jan 14 14:39 (unsaved changes)

File Edit View Insert Cell Kernel Help

Code Cell Toolbar: None

Analyzing the clicklog

Loading the files

```
In [4]: import pandas as pd
        store = pd.HDFStore("/Users/Stian/src/clickpy/mentalhealth_002.h5")

In [5]: store2 = pd.HDFStore("/tmp/mentalhealth_002.h5", complib="zlib", complevel=9)

In [6]: db=store['db']
        store2['db'] = db
        store2.close()

In [65]: lec5 = db[db.lecture_id == 5]

In [66]: lec5['timestamp'] = pd.to_datetime(lec5.timestamp, un

In [67]: lec5.set_index(pd.DatetimeIndex(lec5.timestamp), inpl
```

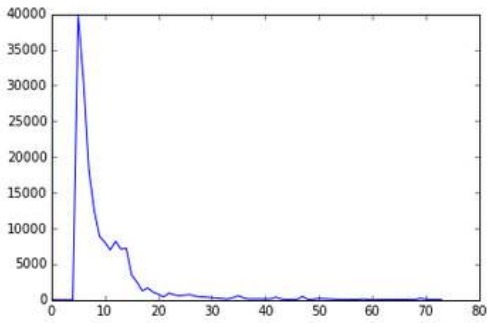
Categorizing video views

```
In [68]: lecnorm = lec5[lec5.playbackRate == 1]
         lecfast = lec5[lec5.playbackRate > 1]
         norm=lecnorm.action.resample("D", how="count")
         fast=lecfast.action.resample("D", how="count")
         tot=lec5.action.resample("D", how="count")
```

Video views

This is how we find everyone watching the video by day

```
In [69]: plot(lec5.action.resample("1D", how="count"))
Out[69]: [<matplotlib.lines.Line2D at 0x11f366490>]
```



Confluence Spaces People Create

Pages / Hatch, Match and Dispatch

Coherence of Research Question and Analysis

Added by Hedieh Najafi, last edited by Hedieh Najafi on Jan 09, 2014 (view change)

Hedieh's note: Here, I am first creating a map that shows the linkage between research objectives, res
that we would use.

Red font color means questions or things that I will have to double check with team members or my no

- Research goals from the proposal:
- Research Questions:
- Data Sources for Each Question and Data Analysis
  - Research Question 1:
  -
  - Research Question 2:
  - Research Question 3:
  - Research Question 4:

~/Dropbox/Laurie/Enc – RStudio

datatable.Rmd | time-segmenting.Rmd | languages.Rmd | Demographic-report.Rmd

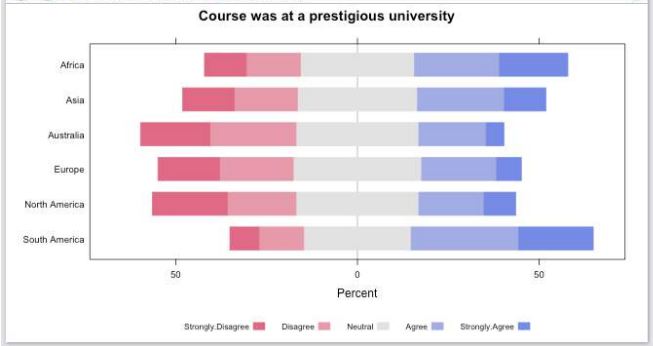
Knit HTML Run Chunks

```
1  # Demographic report
2
3  ## Preparation, turn off output
4  ```{r}
5  library(knitr)
6  opts_chunk$set(warning=F,message=F,error=F,out.width="700px")
7  ```
8
9  Load libraries, set knitr options
10 ```{r}
11 library(ggplot2)
12 library(scales)
13 suppressPackageStartupMessages(require(HH))
14 ```
15
16 Load the data that has been processed with Load-surveys.Rmd
17 ```{r}
18 load("~/Dropbox/Laurie//Enc/db.Rdata")
19 ```
20
21 Let's see which columns we have to play with
```

3:1 (Top Level) R Markdown

Environment History Files Plots Packages

Zoom Export Clear All



Console /Volumes/HD/Bigstuff/Documents/Enc/

```
+   geom_bar(position="fill") +
+   geom_bar(dbgen,mapping=aes("All courses",fill=english),position="f ..."
 ... [TRUNCATED]

> ## ------------------------------------------------------------------------
--
> dbgen <- db[!is.na(db$english), ]

> ggplot(data=dbgen,aes(region,fill=english)) +
+   geom_bar(position="fill") +
+   geom_bar(dbgen,mapping=aes("All regions",fill=english),position="f ..."
 ... [TRUNCATED]
Warning messages:
1: In if (columns == "ALL") { :
  the condition has length > 1 and only the first element will be used
2: In if (xaxis != F) { :
  the condition has length > 1 and only the first element will be used
>
```

Help Git Viewer

R: Group Averages Over Level Combinations of Factors Find in Topic

ave {stats} R Documentation

Group Averages Over Level Combinations of Factors

Description

Subsets of `x[]` are averaged, where each subset consist of those observations with the same factor levels.

Usage

```
ave(x, ..., FUN = mean)
```

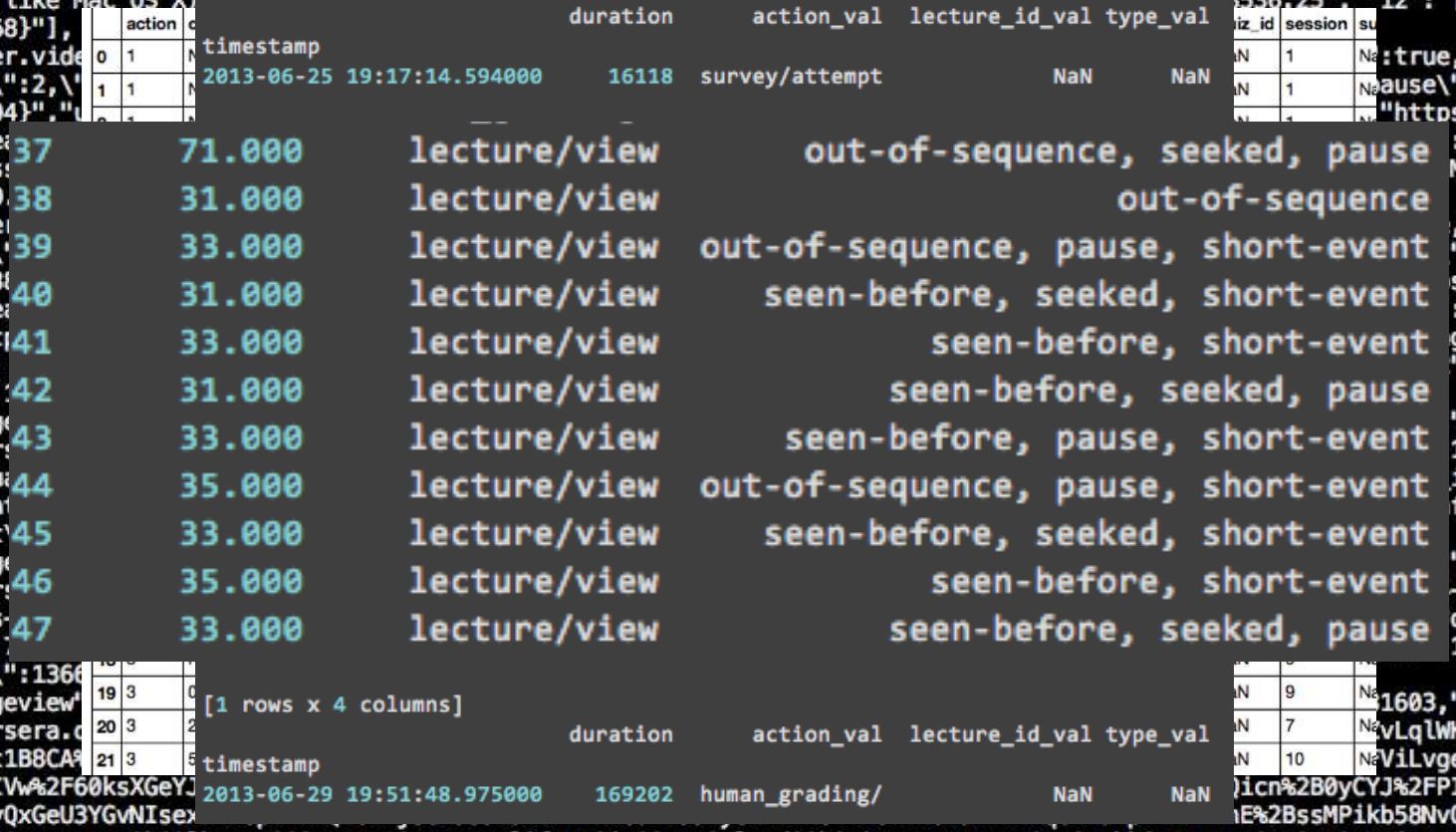
Arguments

`x`	A numeric.
`...`	Grouping variables, typically factors, all of the same `length` as `x`.
`FUN`	Function to apply for each factor level combination.

Value

A numeric vector, say `v` of length `length(x)`. If ... is a1, a2, e.g., `v[i]` is equal to

Clicklog: big data, making it queryable, increasing levels of abstraction

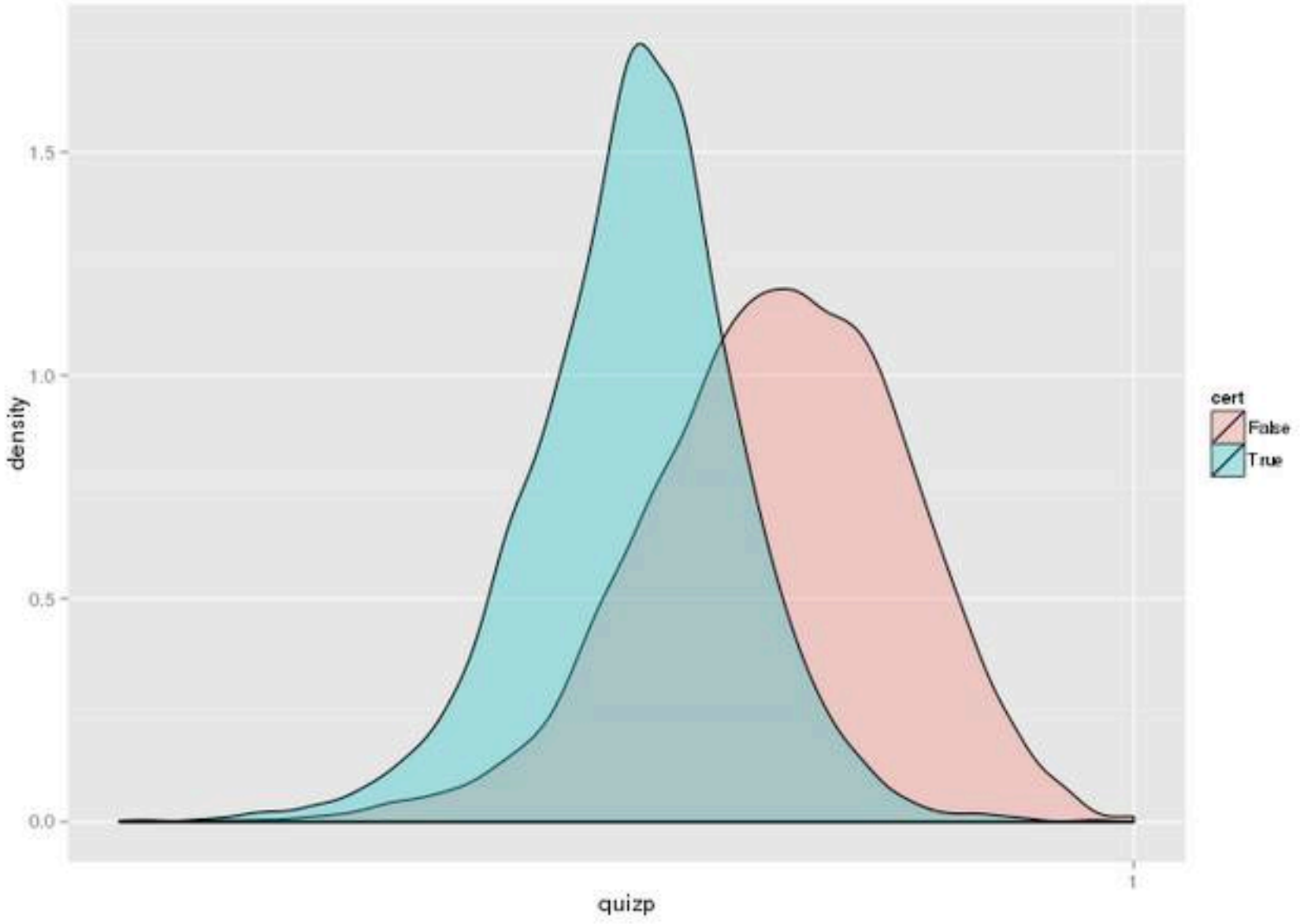


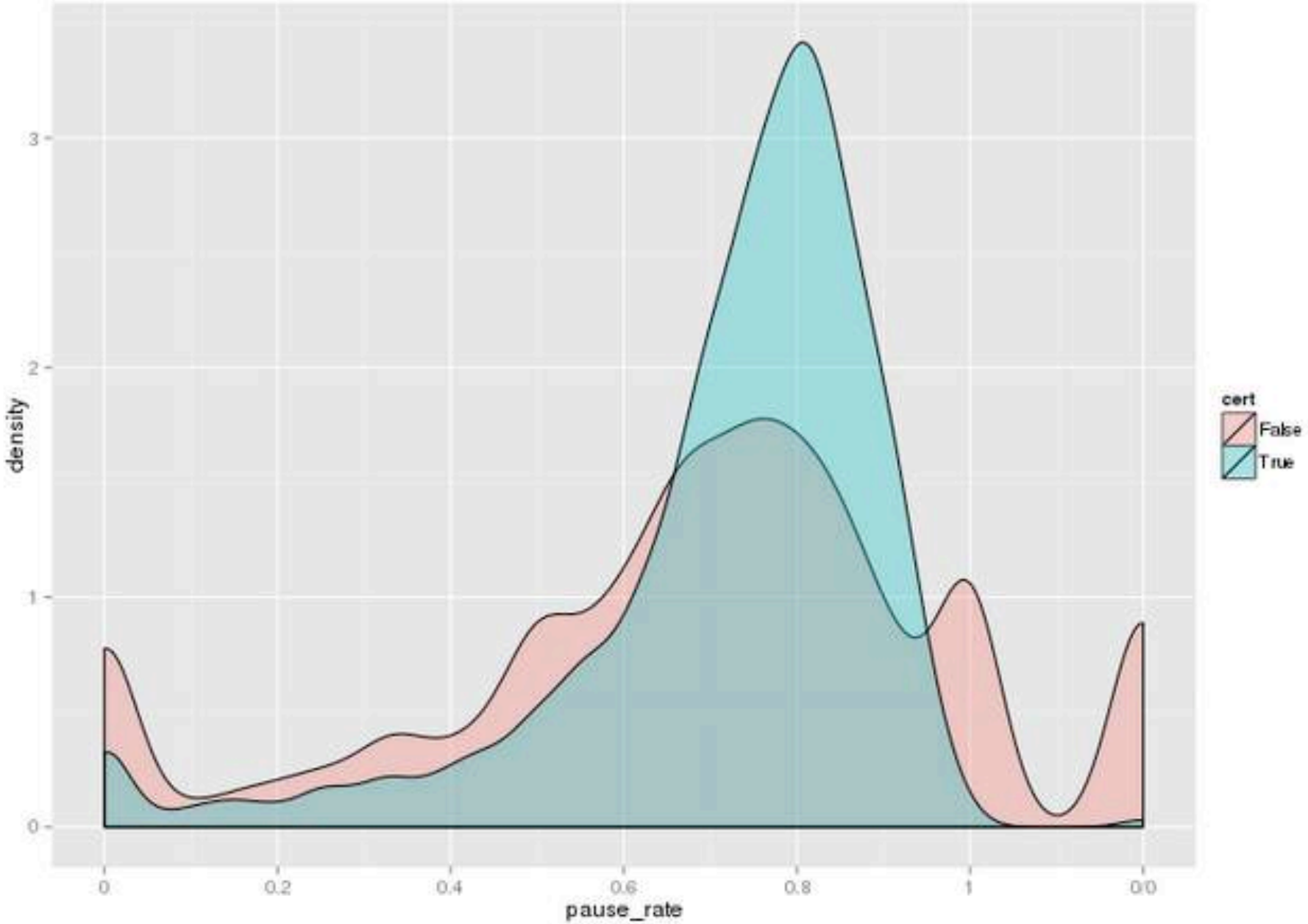
timestamp	duration	action_val	lecture_id_val	type_val
2013-06-25 19:17:14.594000	16118	survey/attempt	NaN	NaN

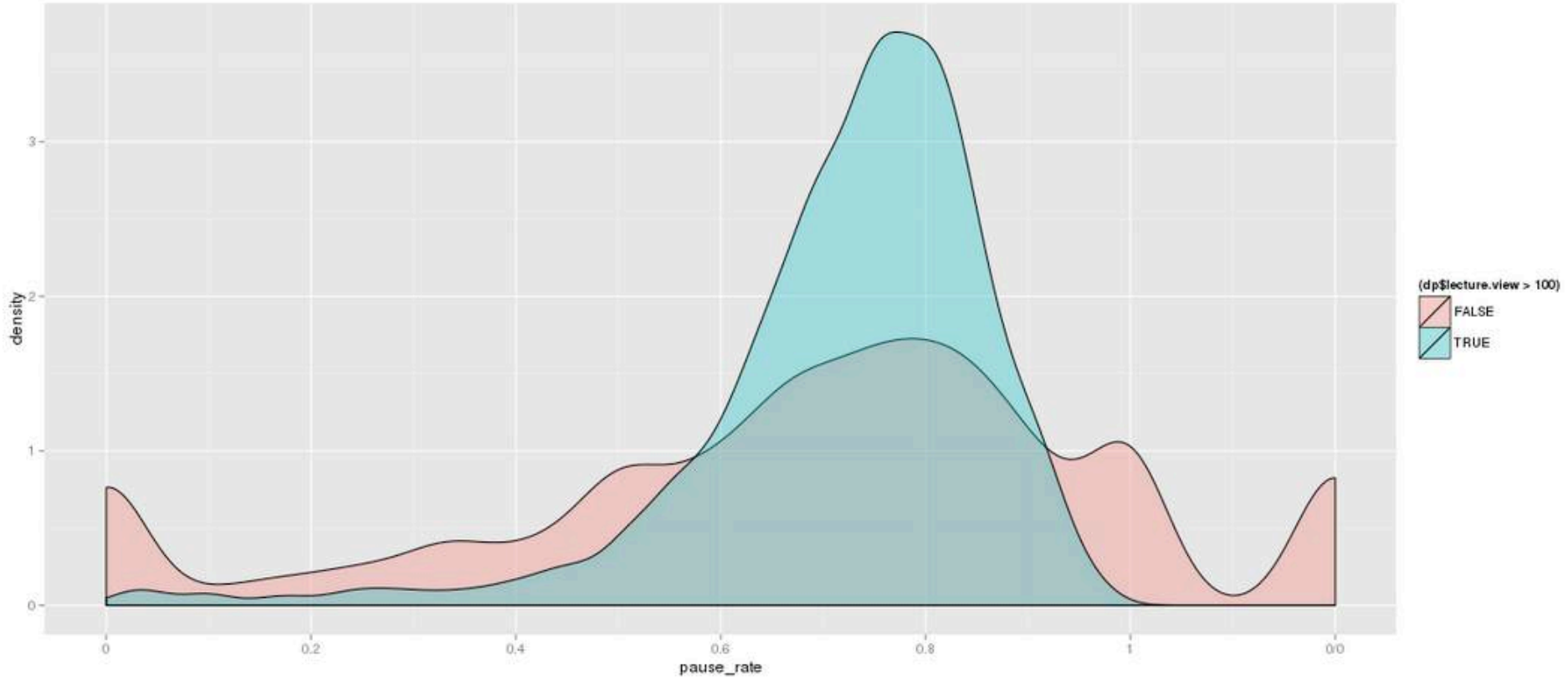
37	71.000	lecture/view	out-of-sequence, seeked, pause
38	31.000	lecture/view	out-of-sequence
39	33.000	lecture/view	out-of-sequence, pause, short-event
40	31.000	lecture/view	seen-before, seeked, short-event
41	33.000	lecture/view	seen-before, short-event
42	31.000	lecture/view	seen-before, seeked, pause
43	33.000	lecture/view	seen-before, pause, short-event
44	35.000	lecture/view	out-of-sequence, pause, short-event
45	33.000	lecture/view	seen-before, seeked, short-event
46	35.000	lecture/view	seen-before, short-event
47	33.000	lecture/view	seen-before, seeked, pause

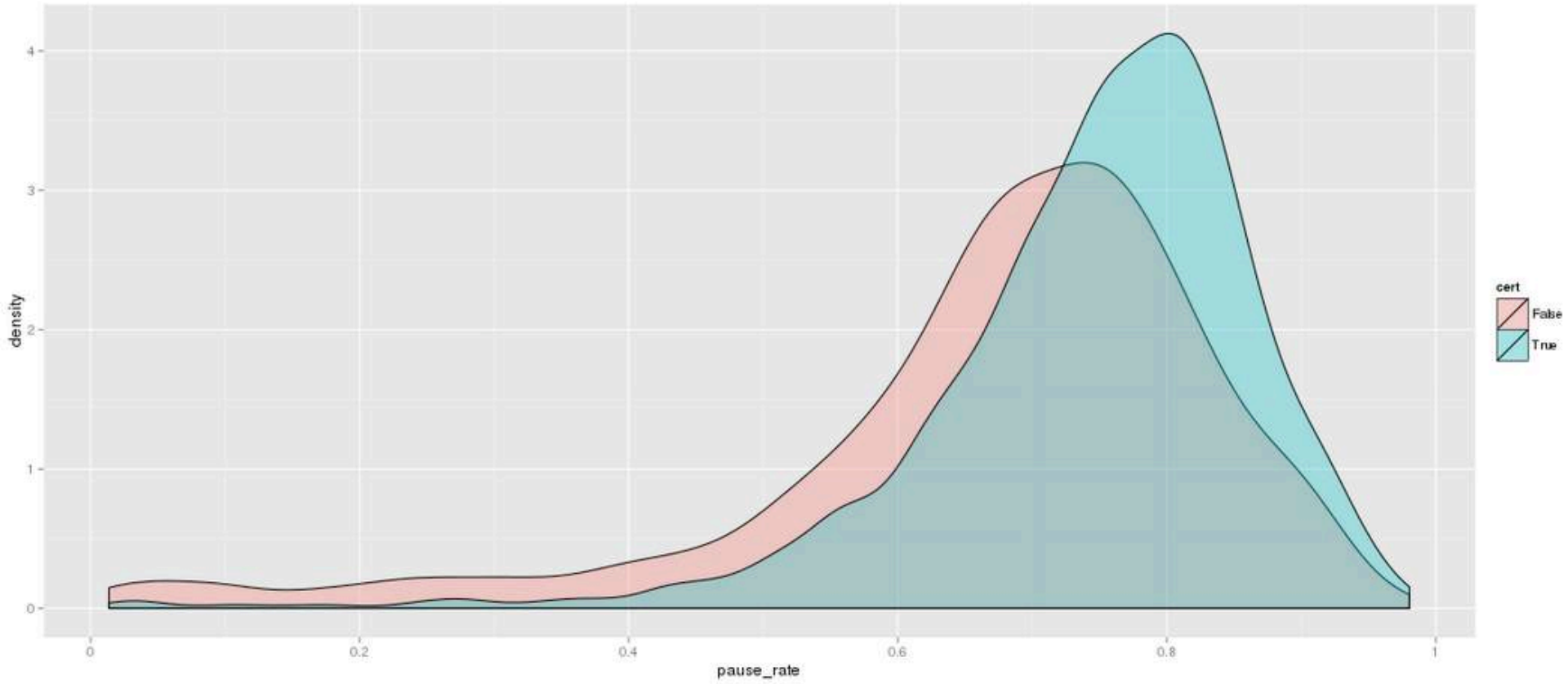
[1 rows x 4 columns]

timestamp	duration	action_val	lecture_id_val	type_val
2013-06-29 19:51:48.975000	169202	human_grading/	NaN	NaN

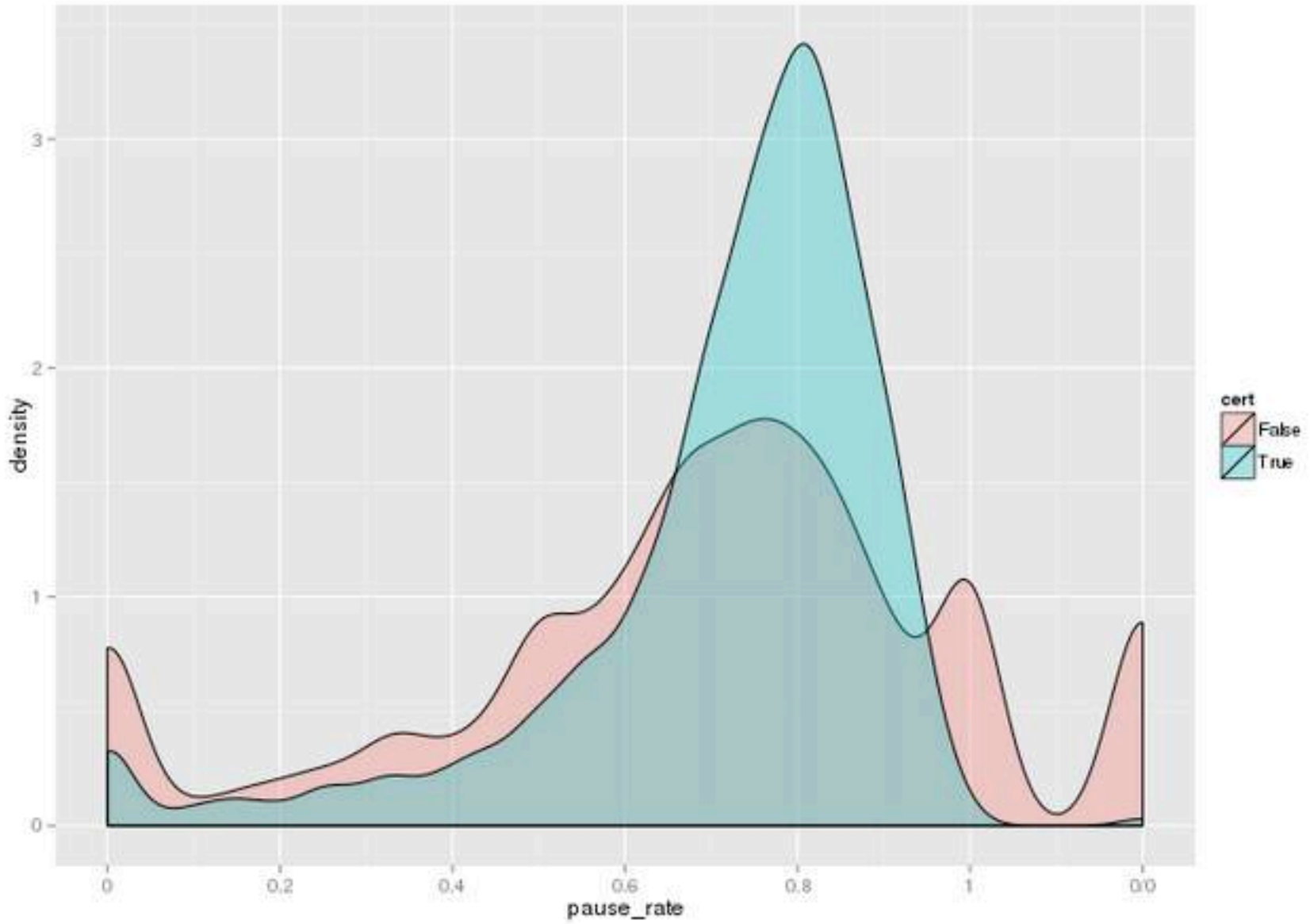

1.5
1.0
0.5
0.0
density
1
quizp
cert
False
True

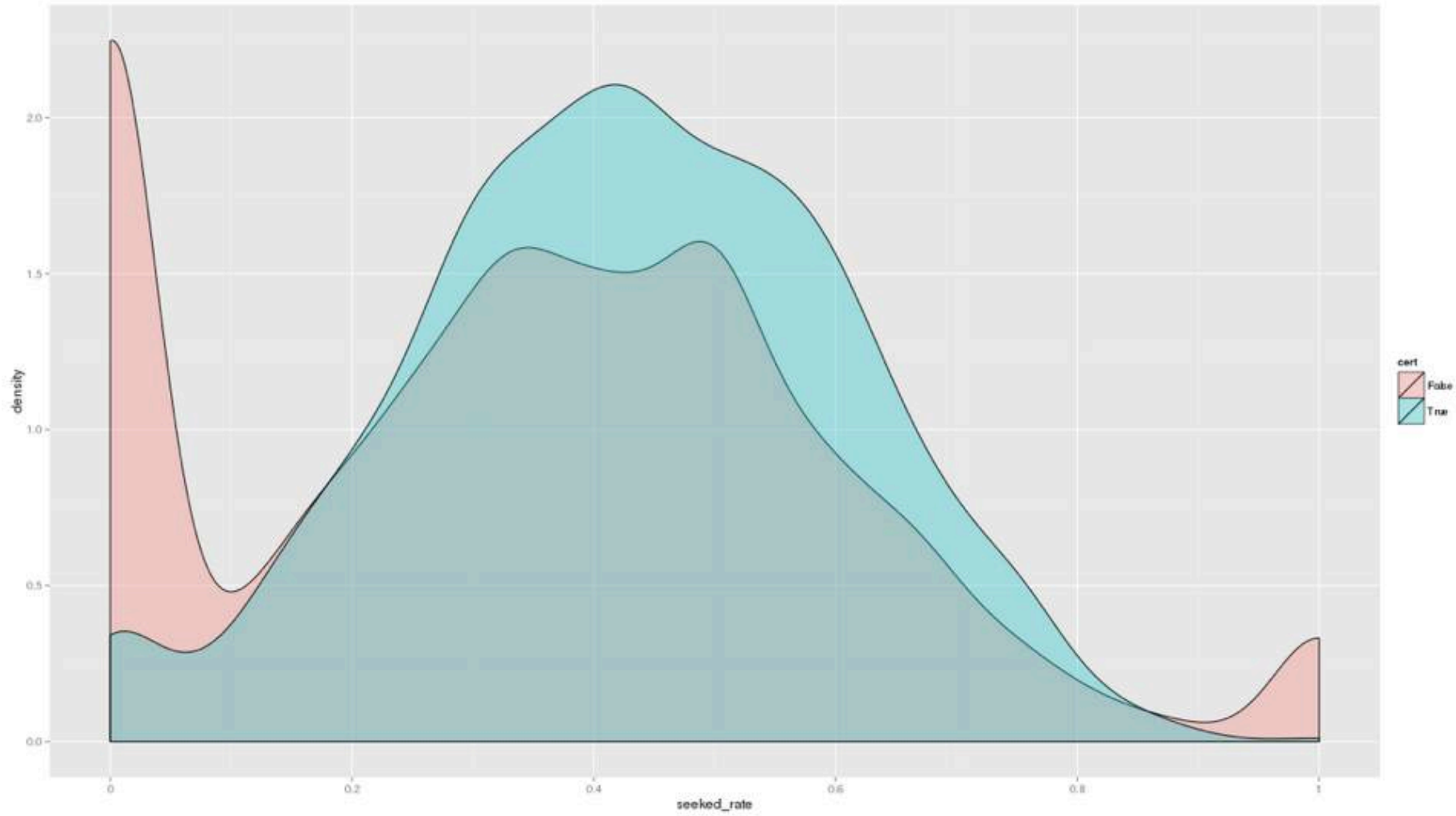

density
3
2
1
0
0
0.2
0.4
0.6
0.8
1
0.0
pause_rate
cert
False
True

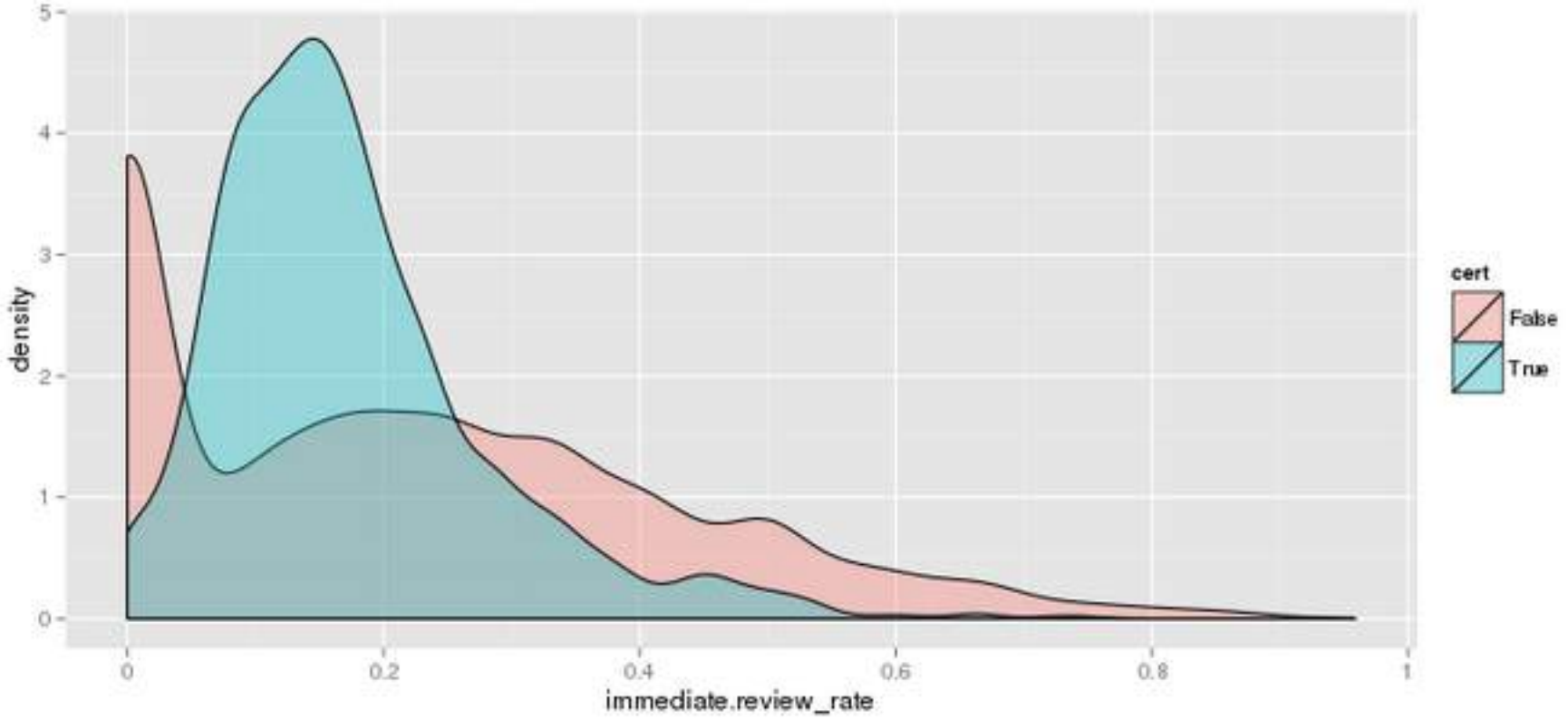

density
3
2
1
0
0
0.2
0.4
0.6
0.8
1
0.0
pause_rate
(dp$lecture.view > 100)
FALSE
TRUE



Students with 100 > lecture.view > 200


density
3
2
1
0
0
0.2
0.4
0.6
0.8
1
0.0
pause_rate
cert
False
True


density
2.0
1.5
1.0
0.5
0.0
0
0.2
0.4
0.6
0.8
1
seeked_rate
cert
False
True


density
5
4
3
2
1
0
0
0.2
0.4
0.6
0.8
1
immediate.review_rate
cert
False
True

sequence	hi-hi-cert-sorted	hi-hi-nocert-sorted	hi-hi-sorted	hi-lo-cert-sorted	hi-lo-nocert-sorted	hi-lo-sorted	lo-hi-cert-sorted	lo-hi-nocert-sorted	lo-hi-sorted	lo-lo-cert-sorted
<{class/index},{lecture/index}>	0.590655885	0.583667422	0.5859498797	0.5816372682	0.5638566957	0.5714839961	0.5619515097	0.6184888037	0.5972074443	0.5653710247
<{class/index},{wiki/view}>	0.5471698113	0.5565626634	0.5534949234	0.5024875622	0.5394021739	0.5235693501	0.4948683623	0.5313136151	0.5175934741	0.5041224971
<{class/index},{lecture/view}>	0.5399820305	0.5217012376	0.5276718117	0.5142469471	0.4575407609	0.4818622696	0.4983886162	0.5223326548	0.5133001064	0.5197290931
<{lecture/index},{lecture/view}>	0.4941599281	0.4734181628	0.4801924996	0.4825870647	0.4385190217	0.4574199806	0.4700282612	0.4850317327	0.4793825017	0.4787985866
<{class/index},{lecture/view,pause}>	0.4858939802	0.4295799198	0.4479722988	0.4595205789		0.4019398642	0.4434032426	0.4430906478	0.4432062123	0.4670200236
<{lecture/view},{lecture/view}>	0.4733153639	0.4191214921	0.4368214097	0.4522840344		0.4065955383	0.4361148297	0.4161477667	0.423662056	0.472614841
<{class/index},{lecture/index},{lecture/view}>	0.4567834681	0.4332403695	0.440929632	0.4364540932		0.409699321	0.4193564381	0.44713208	0.4366728267	0.4331566549
<{lecture/index},{lecture/view,pause}>	0.4486972147	0.4008192435	0.4164563648	0.4332881049			0.4262977837	0.422045264	0.423662056	0.4361012956
<{class/index},{lecture/view,out-of-sequence}>	0.4379155436	0.4102318285	0.4192734315	0.4265038444			0.4156874411	0.4229732966	0.420208695	0.445229682
<{lecture/view},{lecture/view,pause}>	0.4283917341			0.4147444595						0.4284452297
<{lecture/view,pause},{lecture/view}>	0.4273135669			0.4093170511						0.4260895171
<{class/index},{lecture/index},{lecture/view,pause}>	0.4145552561									
<{class/index},{lecture/index},{pause}>	0.4145552561									
<{lecture/index},{lecture/view,out-of-sequence}>	0.4041329739			0.4066033469						0.4175500589
<{class/index},{lecture/view,seeked}>	0.4023360288									
<{class/index},{lecture/view},{lecture/view}>	0.4016172507									
<{lecture/view,out-of-sequence},{lecture/view}>										0.4093050648
<{lecture/view},{lecture/view,out-of-sequence}>										0.410188457

sequence	lo.lo.cert.sorted	lo.lo.nocert.sorted
<{class/index},{forum/index}>	0.2618	0.18585
<{forum/index},{forum/thread}>	0.2506	0.15308
<{class/index},{forum/thread}>	0.2494	0.18696
<{class/index},{forum/index},{forum/thread}>	0.2288	0.13930
<{forum/thread},{forum/thread}>	0.2264	0.14630
<{forum/index},{forum/thread},{forum/thread}>	0.2011	0.10864
<{class/index},{forum/thread},{forum/thread}>	0.1932	0.12875
<{forum/thread},{forum/thread},{forum/thread}>	0.1829	0.10531
<{class/index},{forum/index},{forum/thread},{forum/thread}>	0.1820	0.09776
<{forum/index},{forum/list}>	0.1699	0.10775
<{forum/index},{forum/index}>	0.1673	0.10076
<{wiki/view},{forum/index},{forum/thread}>	0.1670	0.11264
<{forum/index},{forum/thread},{forum/thread},{forum/thread}>	0.1649	0.08009
<{forum/index},{forum/list},{forum/thread}>	0.1572	0.09453
<{class/index},{forum/thread},{forum/thread},{forum/thread}>	0.1572	0.09309
<{forum/thread},{forum/index}>	0.1555	0.09842
<{class/index},{wiki/view},{forum/index},{forum/thread}>	0.1549	0.10409
<{class/index},{forum/index},{forum/list}>	0.1537	0.10009
<{class/index},{forum/index},{forum/index}>	0.1519	0.09042
<{forum/thread},{forum/thread},{forum/thread},{forum/thread}>	0.1493	0.07876
<{forum/index},{forum/index},{forum/thread}>	0.1487	0.08054
<{forum/index},{forum/thread},{forum/index}>	0.1472	0.08354
<{class/index},{forum/index},{forum/list},{forum/thread}>	0.1422	0.08831
<{forum/thread},{forum/index},{forum/thread}>	0.1419	0.08087
<{lecture/view},{forum/index},{forum/thread}>	0.1402	0.08787
<{class/index},{forum/thread},{forum/index}>	0.1390	0.08731
<{forum/index},{forum/thread},{wiki/view}>	0.1381	0.08509
<{lecture/view,pause},{forum/thread}>	0.1378	0.08776

Next Steps

- Examine outcome measures
- Repeat with other MOOCs
- Refine feature identification