Educational Data Mining Approaches for Digital Libraries
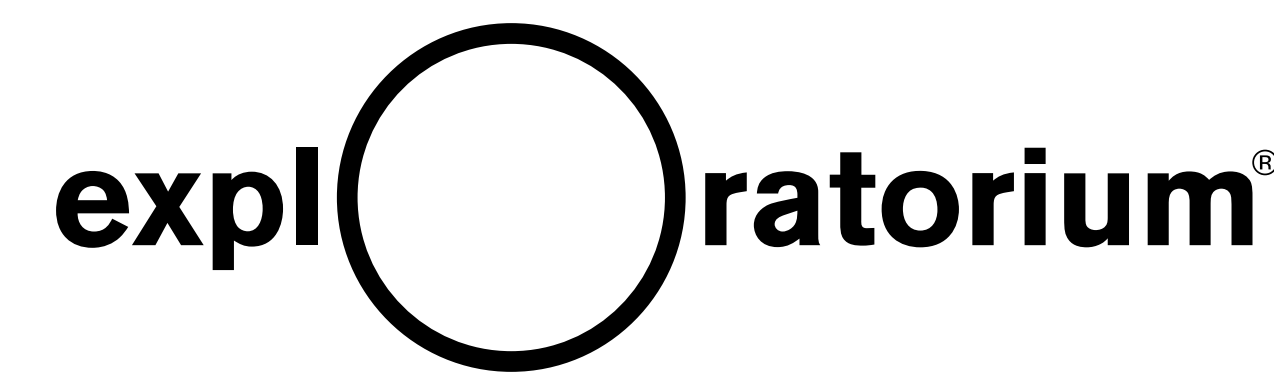
Mimi Recker, Beijie Xu, Bart Palmer *Utah State University*
Sherry Hsi, Rob Rothfarb *Exploratorium*
Project Site: http://edm.usu.edu

UtahState University · explOratorium · NSDL · NSF

RESEARCH GOALS

- Collect Web usage data in the Instructional Architect and the Exploratorium's Learning Resources Collection and apply Web metrics and data mining techniques to infer teacher behaviors and learning online.

- Contribute new knowledge regarding Knowledge Discovery from Data (KDD) and the application of Educational Data Mining (EDM) to digital libraries.

DATA SOURCES

Instructional Architect (IA)	Exploratorium LRC (ELRC)	Other Data Sources
Web usage: Web server log Google Analytics IA relational database	*Web usage:* Web server log Google Analytics	U.S. Census Data NCES Demographic Data
Other: Registration profile Online survey	*Other:* Online survey Search term capture Interviews *(tbd)*	

WEB METRICS

| Metric	Definition
Visits	Sequence of user transactions from a single IP within a certain time period (considered a more robust measure than "hits")
Unique visitors	A measure of a site's true audience size
Page views	Number of times Web pages are accessed during a single visit, including repeated viewing of the same page
Visitor origin	Country or region of visitor origin
Operating Sys/Browsers	Operating system and browser used by visitors
Length of visit	Time spent by visitor on site
Entry/Exit Pages	First and last pages accessed by visitors
Referrer	Last page the user visits before landing on the site
Bounce rate	The percentage of visitors who "bounce" away to a different site

EXAMPLE 1

IA Monthly Visits
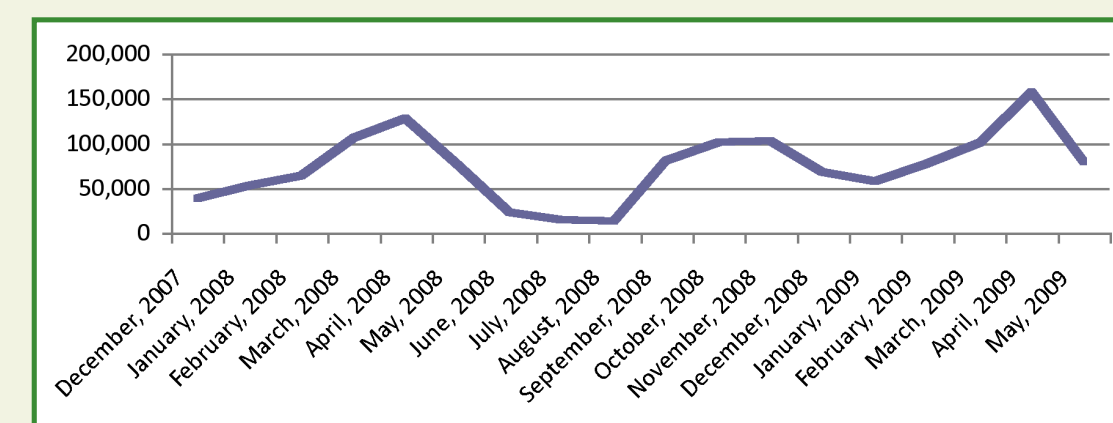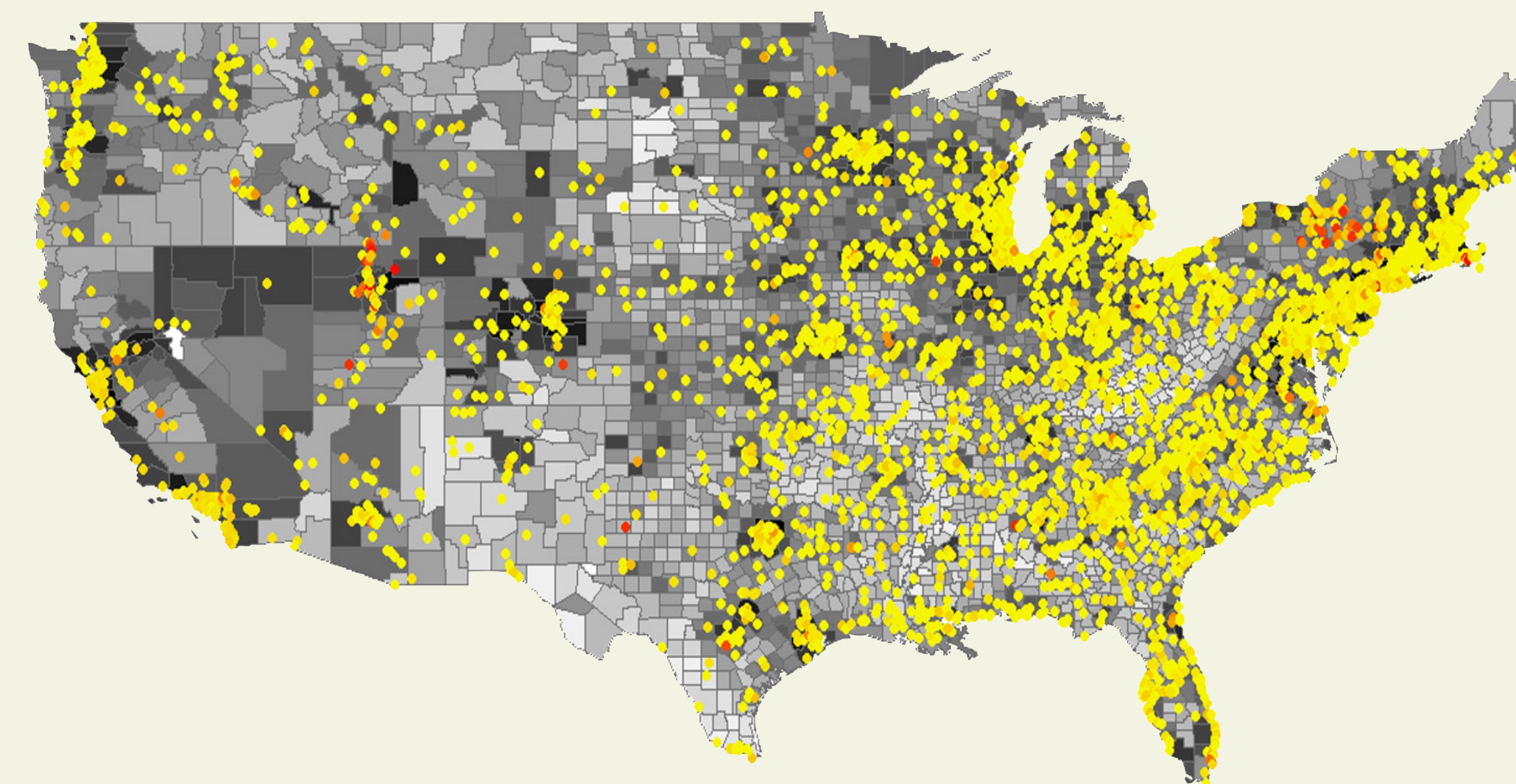

ELRC Monthly Visits
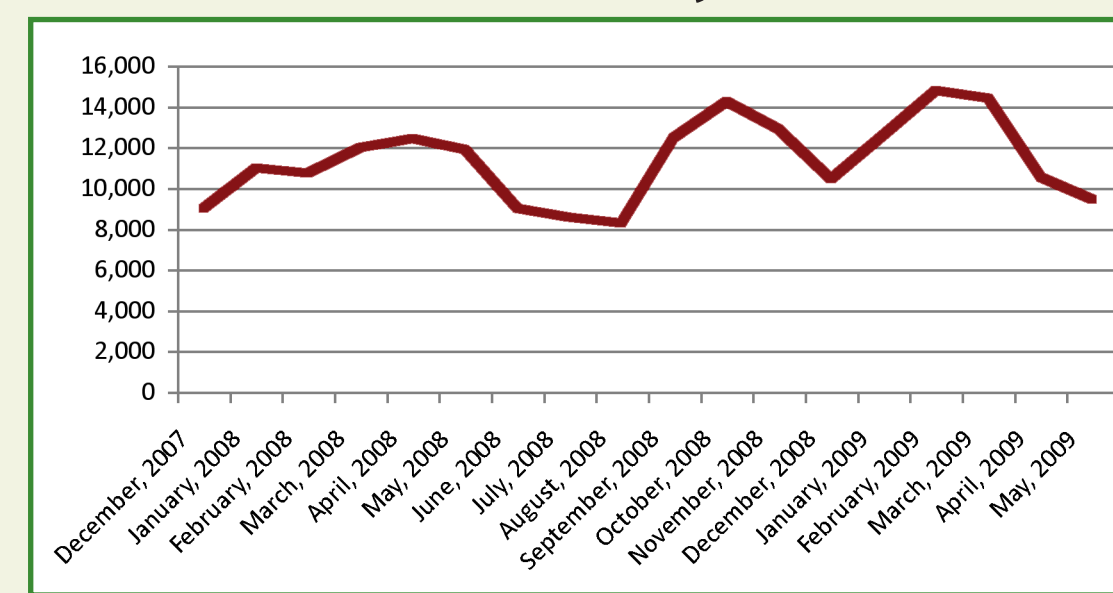
EXAMPLE 2

Visits from the Contiguous United States

(a) Instructional Architect   (b) ELRC
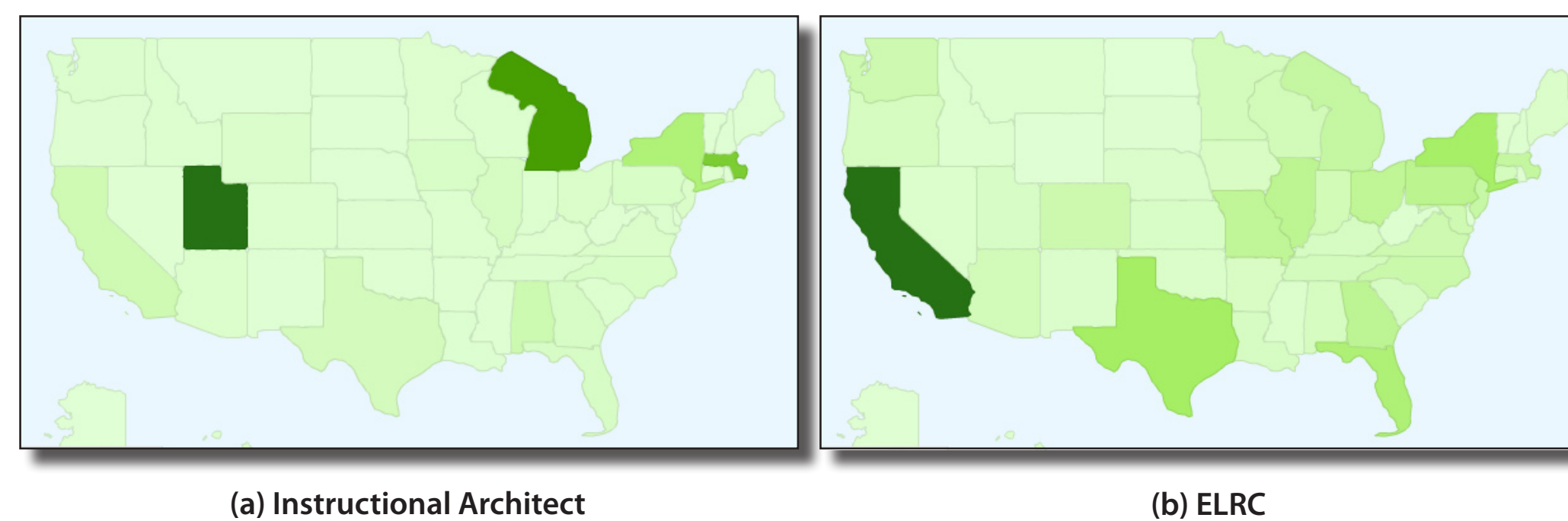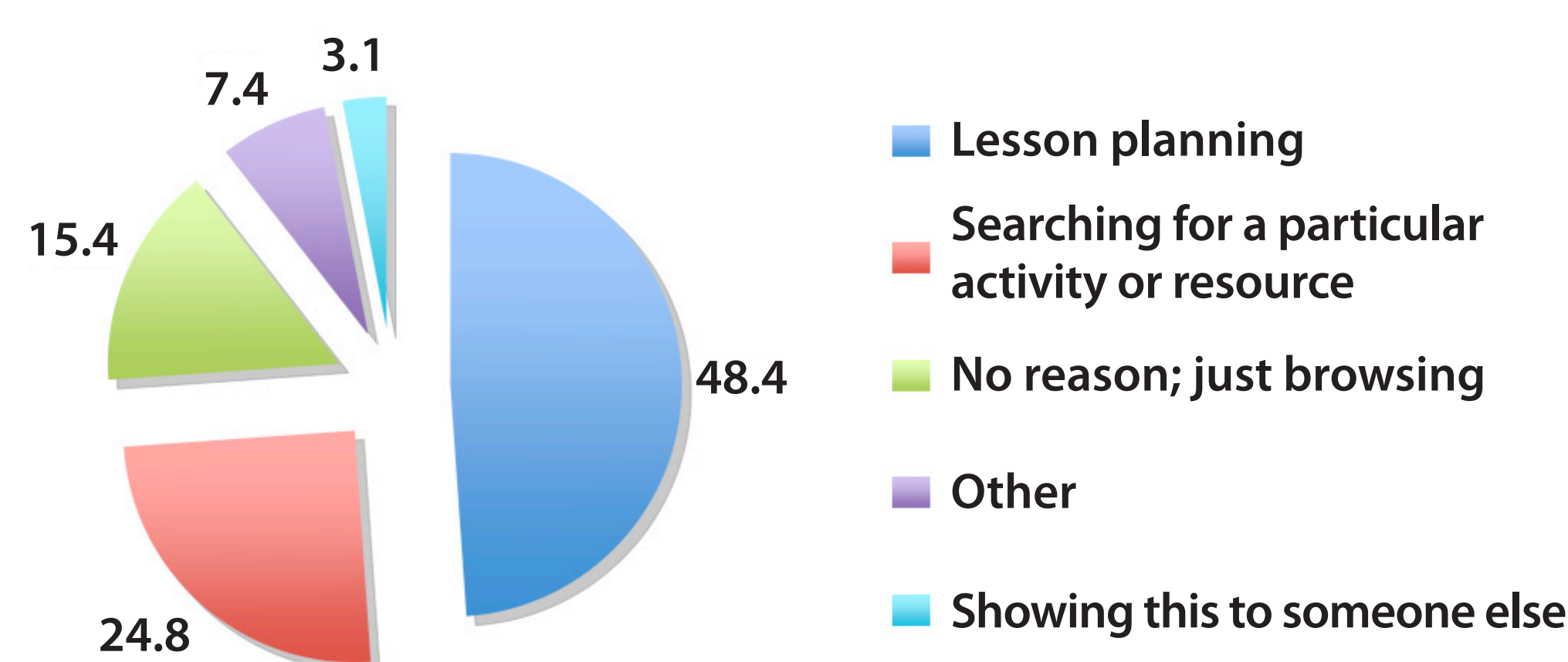
EXAMPLE 3

GeoAnalysis
Visits Overlaid on Median Family Income (IA)

EXAMPLE 4

Combining with Pop-up Survey in Application (ELRC)
Visitor Purpose (July–Sept 2009, N = 714)

- 48.4 ■ Lesson planning
- 24.8 ■ Searching for a particular activity or resource
- 15.4 ■ No reason; just browsing
- 7.4 ■ Other
- 3.1 ■ Showing this to someone else

EDUCATIONAL DATA MINING

Knowledge Discovery from Data /Ed Data Mining Process

Phase I   Data Preprocessing
- Data cleaning and path completion
- Data integration
- Data selection
- Data transformation

Phase II   Applying Data Mining Algorithms

Phase III   Interpretation and Post-processing; Evaluation and presentation

DATA MINING APPROACHES

Approach		Application for the IA	
Algorithm	**Usage**	**IA data**	**Intention**
Clustering	Cluster data into groups	Use of resources Use of projects Online behaviors	Identify teacher groups
Association rule	Examine the relations among variables	Use of resources Use of projects Online behaviors	Identify the relations among teacher-related features
Sequential pattern	Examine the pattern in time-series transactions	Click-stream data	Identify the IA pages often accessed together

LATENT CLASS ANALYSIS— CLUSTERING RESULTS

		Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Cluster Size		0.29	0.24	0.21	0.16	0.10
Indicators	**Range**					
percent_student_projects	0–1	0.96	0	0.34	1	0.44
percent_private_projects	0–1	0	1	0.29	0	0.32
percent_copy_projects	0–1	0	0.15	0.30	0	0.32
percent_public_projects	0–1	0.96	0	0.40	0	0.43
Visit stickiness	0.22–8.93	1.16	0.62	0.99	0.71	2.98
Project stickiness	0.15–8.06	1.99	0.20	0.66	0.32	2.12
Resource stickiness	0–29.91	1.06	0.39	0.77	0.57	4.22
Browse others	0–1	0.44	0.44	0.20	0.59	0.21

Cluster 1: Goal-oriented, Willing-to-share
Cluster 2: Inactive, Less motivated
Cluster 3: Lukewarm performers
Cluster 4: Goal-oriented, Less willing-to-share, Less devoted
Cluster 5: Productive, Active, Adapters