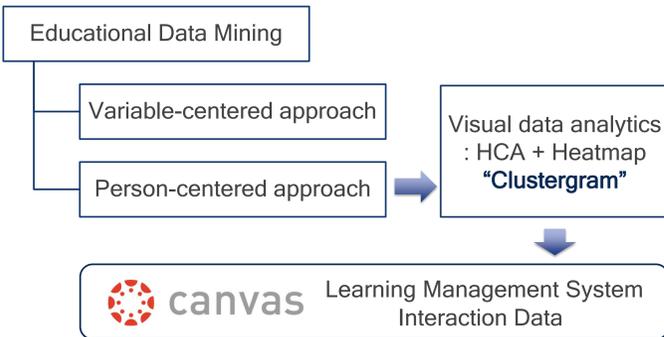


# Hierarchical Cluster Analysis Heatmaps and Pattern Analysis: An Approach for Visualizing Learning Management System Interaction Data

Ji Eun Lee, Utah State University  
 Mimi Recker, Utah State University  
 Alex Bowers, Columbia University  
 Min Yuan, University of Utah

## I. INTRODUCTION



## II. RESEARCH QUESTIONS

- To what extent do clustergrams help understand patterns of student activity in the course?
- How do these patterns of activity relate to student learning outcomes?

## III. BACKGROUND

### Educational Data Mining (EDM) and Learning Management System (LMS)

- Educational data mining applies computational and statistical techniques to find patterns in the large datasets automatically collected by learning technologies, here specifically LMS.
- Prior EDM research have taken a variable-centered approach by examining usage at an aggregated level.
- In contrast, we take a person-centered approach to visually investigate what sub-groups of students may share common patterns, and how these relate to their learning outcomes.

### Hierarchical Cluster Analysis (HCA) Heatmaps

- HCA classifies related units in an analysis across high dimensionality data, and can be combined with heatmap visualizations to create a clustergram (Bowers, 2010).
- When combined with the pattern analysis from HCA, a heatmap is easily interpreted by the human eye as representing the overall intensity patterns of the participants across the dataset.

## IV. DATA SOURCES and METHODS

- Student interaction data recorded by Canvas LMS
- Freshman-level online mathematics course taught during the fall 2014 semester (N = 139)

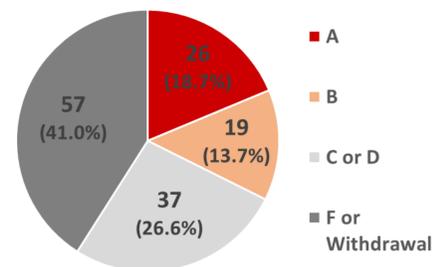


Figure 1. Final grade distribution

- HCA: Euclidean distance measure with the average method

Clustergram using the semester summary data	Clustergram using weekly summary data
Data were transformed to z-scores	Used raw data
HCA applied to rows and columns	HCA applied to only the rows

```
> Heatmap(mat, name = "Z-score",
column_title = "Canvas features",
column_hclust_side = "bottom",
cluster_rows = TRUE, clustering_distance_rows =
"euclidean", clustering_method_rows = "complete",
row_names_gp = gpar(fontsize = 7),
cluster_columns = TRUE,
clustering_distance_columns = "euclidean",
clustering_method_columns = "complete",
col = colorRamp2(c(-3, 0, 3),
c("blue", "white", "red")),
km = 3) + ha
```

Figure 2. Example of the R code used for applying HCA and k-means clustering

## V. RESULTS

### 1 CLUSTERGRAM USING SEMESTER SUMMARY

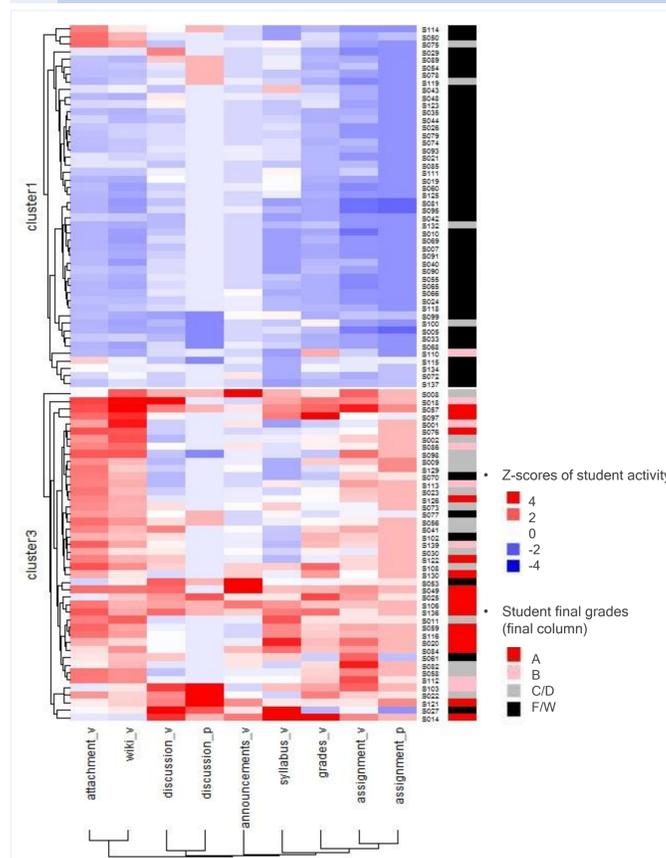


Figure 3. Section of the clustergram using the semester summary data

Table 1. Correlations between student Canvas activity and final grades (N = 139)

	1	2	3	4	5	6	7	8	9
1.final grades									
2.announcements_v	.16								
3.assignment_v	.61*	.30*	.78*						
4.assignment_p	.70*	.25*	.57*	.56*					
5.grades_v	.56*	.20*	.44*	.44*	.35*				
6.attachmen-t_v	.35*	.09	.46*	.36*	.50*	.14			
7.syllabus_v	.35*	.27	.34*	.21*	.35*	.27*	.43*		
8.discussion_v	.20*	.27	.19*	.12	.23*	.09	.17*		
9.discussion_p	.11	.17	.55*	.48*	.40*	.86*	.25*	.45*	
10.wiki_v	.42*	.27	.78*	.56*	.35*	.14	.43*	.37*	.13

- Most students with lower activity (Cluster 1) either received a grade of F or withdrew (W) from the course, whereas many students with higher activity (Cluster 3) received a grade of A (Figure 3).
- Excluding the assignment variables, students' views of their own grades (grades\_v) were most highly correlated with their final grades (Table 1). This aligns with the findings from previous studies showing that student monitoring of their performance had a positive impact on achievement (Lan, 1996).
- The students with a grade of A (sub-cluster B) tended to show higher attachment views at the beginning of the course and more consistently throughout the semester. However, the students with grades of C/D (sub-cluster C) tended to have higher attachment views at the end of the course, representing perhaps a less-successful 'cramming' strategy (Figure 4).

## VI. SIGNIFICANCE

- Combining the cluster analysis heatmap visualizations with the large dataset provides a unique opportunity to examine the patterns of student activity and LMS features as they relate to overall student outcomes.
- This type of visual data analytics expands the number of tools available for instructors and administrators to help identify the features and specific LMS interaction data that are most useful to their students.

### 2 CLUSTERGRAM USING WEEKLY SUMMARY

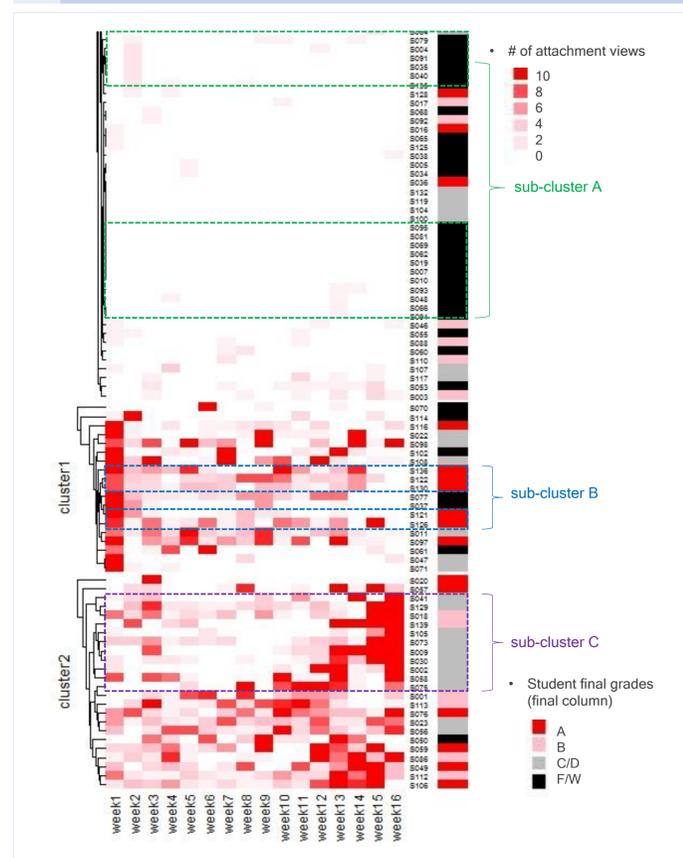


Figure 4. Section of the clustergram for the number of students' attachment views by week