Analysis of student usage of online videos in synchronous and asynchronous mathematics courses

by

Melea R. Roman

B.S., Cedar Crest College, 2019

A THESIS

submitted in partial fulfillment of the
requirements for the degree

MASTER OF SCIENCE

Department of Mathematics
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2022

Approved by:

Major Professor
Dr. Andrew Bennett

Copyright

Abstract

The push toward online learning in response to the Covid-19 pandemic has provided an opportunity to evaluate student behavior and outcomes in asynchronous and synchronous mathematics courses with an unprecedented population of students. In this study, data is collected from Canvas tracking student utilization of lecture videos. This data is analyzed under the lens of principal component analysis and partitioning around medoids to cluster students into behavioral groups. This data is then compared with grade outcomes to provide insight into possible behavioral best-practice for asynchronous and synchronous online learning in Elementary Differential Equations and History of Mathematics courses.

Table of Contents

List of Figures

List of Tables

Acknowledgments

I would like to thank my committee members Dr. Craig Spencer and Dr. Rekha Natarajan who dedicated their time to seeing me through this process despite being coordinators for some of the largest courses the math department has to offer. I would also like to thank my advisor Dr. Andrew Bennett who gave me guidance for not only this thesis but for also my job search as I prepared to transition out of graduate school. Finally, I want to thank all of my peers in the math department who have given me their friendship and support over the past two years.

Dedication

I dedicate this thesis to my younger siblings Dillon, Julia, Gabriel, Markus, Mary, Farrah, Alice and Calypso in hopes that they are emboldened to keep working hard toward their goals.

Chapter 1

Introduction

This paper examines student online behavior and outcomes in an online synchronous History of Mathematics course and an online asynchronous Elementary Differential Equations course both of which took place between the 2019-2020 and 2020-2021 academic years at Kansas State University in the midst of the COVID-19 pandemic.

1.1 Motivation

Mathematics education research is well established at the K-12 and remedial college level. Research in recent years has delved into higher level mathematics courses at the university level. This study is interested in Elementary Differential Equations and History of Mathematics courses.

Why might we be interested in these courses? Elementary Differential Equations is an essential course for most engineering and mathematics majors. This course uses student's calculus 1 and calculus 2 knowledge to relate unknown variables with their derivatives. It has been said that "mathematics [is] the language of science, and differential equations are one of the most important parts of this language as far as science and engineering are concerned".

History of Math is a course that gives students a deeper developmental understanding of the math they have studies with hands-on practice and reading translations of primary texts. It is often required for education and mathematics majors. While less popular than Elementary Differential Equations and far less studied, it serves a group of students set to teach mathematics to the future generations of math students.

Cognitive learning styles, a complex and widely studied topic in education and psychology, play a critical role in a student's mathematics experience. This is often defined as the student behavior in organizing and processing information in order to learn [2]. These cognitive styles have been re-imagined for mathematics education with early research citing purpose and process as the defining components of mathematics learning [3]. Further studies began to expand these cognitive styles for specific courses. These studies cite that a students ability to discern what information is important plays a critical role in student calculus success [4]. These studies, in conjunction with more recent research, has shifted focus on how to recognize patterns in student learning behavior in order to better teaching practices and stage learning interventions for struggling students.

Research looking at online education has primarily focused on video content as gauge for student success, but recent studies have begun to analyze different aspects of online behavior in assessing student learning and success. What makes current research different from past research in online learning is the change in student population. Online courses have been offered as a flexible alternative to traditional courses for students with other job, family or course obligations. Students have particular motivations for taking an online courses, whereas during the COVID-19 pandemic nearly the entire student population was forced to take courses in an online setting whether they wanted to or not. This gives an additional opportunity to examine online learning behavior for a large, more general population of students that is not usually available.

1.2 Limitations

This study is limited to the pool of students from Kansas State University enrolled in 2 mathematics courses. Further, this was during the 2020-2021 academic year in the middle of the COVID-19 pandemic. These conditions not only limit the population scope, but also limit the ability to reproduce conditions of the study.

Further limitations to the study include a lack of direct student input about their behaviors; therefore, there are many aspects about these behaviors that we are unable to discern. For the asynchronous course, we do not have a way of knowing how they watched their videos. It is reasonable to assume that some students may not pay close attention to the videos while they are playing, but still show up as having “watched“ the video. Students that clock-in low view times could reasonably be watching videos in a group on a single account. They have watched the video, but in our data set it says they have not. These are just a few of many possible explanations for perceived online behaviors.

Similarly, for the synchronous online course, we do not have any data on the live video participation. This means that any logged viewings of posted videos are automatically “review“ and do not say anything about behaviors around lecture viewing or participation.

Other limitations come from the population pool. For the MATH 340 differential equations course the students have low diversity in terms of general educational background. A large majority of the students are specifically engineering and mathematics students. Mathematics training up until this point have been largely the same, and study habits may not reflect a wider student body.

1.3 Definitions of Key Terms

The following terms will be used in this thesis to describe different classes, assignments and actions.

Definition 1 (Synchronous). [5] Synchronous-driven online courses are those in which the instructor and students meet regularly online in real-time meetings for lectures or other activities. The course may also be supplemented with other online activities or materials, but the majority of the student seat time is offered through synchronous meetings.

Definition 2 (Asynchronous). [5] Asynchronous-driven online courses are those in which students are rarely if ever, required to meet in a synchronous setting and instead complete asynchronous learning activities and engagement on their own time. Some synchronous online office hours or group meetings may be included in this type of course, but the majority of materials will be available to students through the Canvas course site.

Definition 3 (Lecture)**.** [6] A lecture is the traditional method of instruction in which students are taught a subject by a member of the faculty. Typically instruction delivered via lecture tends to be more instructor centered than discussion oriented.

Definition 4 (Recitation)**.** [6] Laboratory instruction is the application of methods and principles to student-oriented practice, often in a hands-on way. This designation is primarily used in science and engineering. The laboratory time is separate from the lecture although it is often associated with a lecture component.

Definition 5 (Labs)**.** [6] A recitation is an interactive meeting that combines formal presentation, review and interaction between the students and a faculty member. It is usually combined with a lecture as the primary component. Recitation sessions often review the lecture content, expand on the concepts and usually allow for question and answer time.

Definition 6 (Coverage)**.** In regard to student watching behaviors, coverage tells us how much a of a video a student has watched in terms of the percentage watched.

Definition 7 (Watched)**.** Did a student click play on a video.

Definition 8 (Reviewed)**.** A student watched a video again at least one day after originally watching the video.

1.4 Background

1.4.1 Literature Review

There have been several recent papers outlining instruction and online instruction of differential equations. These studies focus on class structure and student outcome as opposed to student behavior and outcome. A 2021 paper at the University of Montreal found that over the COVID pandemic their online sections of differential equations actually out-performed their in-person sections[7]. However, they offer little information on the details of the structure of the course.

When it comes to student outcomes in terms of active and passive learning, studies have generally supported active learning when considering student satisfaction; however, these studies have primarily been done for in-person instruction. A 2009 empirical study on active versus passive learning looked specifically at student outcomes when comparing the teaching/learning styles as opposed to student satisfaction.[8] They found, in agreement with a majority of published research on the subject, that the two styles do not significantly affect student outcomes.

Technological advances and increased access to internet has edged virtual education into mainstream over the past decade. The COVID-19 pandemic thrust virtual education as the main source of education for millions of students mid-year without preparation. From this, many studies have come out in an attempt to assess student learning and outcomes to better serve students in a virtual setting.

Students in asynchronous courses are forced to rely on strong self-regulated learning behaviors. As defined by the US Department of Education's Literacy Information and Communication System, self-regulated learning is "one's ability to understand and control one's learning environment.[9]. Self regulating abilities include goal setting, self monitoring, self-instruction, and self-reinforcement". These skills, which may not be fully developed in students at certain levels, are at the heart of asynchronous online learning where a teachers in-class guidance is limited. Online instruction brings the challenge of student engagement being hard to assess and enforce. However, it does bring a unique opportunity for data analysis. Student online activity, specifically online activity in online learning platforms creates a lot of data.

Video watching data can tell us how long a student watched a video, how many times they pressed pause, play, if they sped up a video or even if they skipped around. Recent studies have looked into this data on video engagement as an outcome predictor. We call this "click-based" data, meaning data created from a student clicking different buttons in an online platform. Several recent studies have look at click-based performance prediction. On a large scale, this 2021 study assessed online open-learning courses EdX and Coursera.[10] These are courses open to a huge online audience that are not connected to a specific institution. Because of this, students come from a wide set of backgrounds and ages. This study worked to create a model for tracking student click activity to provide a quantifiable way to assign a level of student participation in a given online environment.

This idea has been put to use in other studies. One recent 2021 study looked to answer questions on how video engagement, using a click-based model, could cluster student behaviors and predict outcomes. They classified video click-based engagement as "manipulating" behaviors as pausing, fast-forwarding, and rewinding.[8] They then used principal component analysis and clustering using partitioning around medoids to analyze the data finding that the important behavioral patterns included rates of "pause" and "play" and peer-to-peer interactions using "view comment" and "comment" options on videos. The clustering put the students into two groups with distinguishing factor being the social interaction aspect of PC2. In particular, the first cluster contained "active learners" where the second cluster contained "passive learners". The first cluster performed better on exams than the second, which somewhat contradicts previous studies comparing active and passive learners.

Synchronous learning offers much of the traditional structure of classroom learning with a set schedule and assignment timeline. A 2021 study looked at student behavior specifically in synchronous online behavior by assessing participation levels to gauge student engagement over the course of the semester.[11] This study focused on college students from a variety of majors and primarily analyzed student participation by the attendance log. The results did not compare against student outcomes but did find trends of decreased participation as the semester progressed. They also found that participation patterns did vary depending on what time of the day the course took place.

1.4.2 Research Questions

I pose the following questions:

1. Can we identify patterns in student behavior by their online utilization of lecture

videos?

2. Are there specific behavioral patterns from students who succeed vs. do not succeed in the course?

1.4.3 Course Description MATH 570 History of Mathematics

MATH 570 is History of Mathematics: "A survey of the development of mathematics from ancient to modern times." [12] This course is a 3 credit course with a MATH 220 Calculus I prerequisite. It is typically taken during the spring semester. It is a requirement for students majoring in secondary education with an endorsement in mathematics. It is also a popular elective option for general mathematics majors and students looking to receive a Primary Text Certificate.

The course was organized online through Canvas using modules. Modules were organized by topic in the chronological order they appeared in lectures and included pdf documents for readings and homework, links for lecture recordings and participation quizzes.

The data used in this study was taken from the spring 2021 semester. The course was run synchronously with zoom lectures. Prior to lectures, reading assignments would be assigned and a small quiz would be opened for a limited time as a mark of participation. These quizzes opened on Canvas, and lecture participation was not required to take them. Lecture videos were posted and made available for viewing for the rest of the semester following its original broadcast. Lecture videos from past semesters were also made available. This course does not have recitations or labs guided by a graduate student. This course also did not have an assigned textbook, but primary text excerpts were provided to students through Canvas.

Assignment Type	Description
Written Homework	12 assignments. Roughly, one assignment per week
Exams	1 midterms and 1 final
Quizzes	Due at the beginning of each class, usually required short readings
Papers	3 over the course of the semester based on conceptual development of mathematics with note historical and cultural influences

Table 1.1: MATH 570 Graded Assignments

1.4.4 Course Description MATH 340 Elementary Differential Equations

MATH 340 is Elementary Differential Equations: "Elementary techniques for solving ordinary differential equations and applications to solutions of problems in science and engineering."[13] This course is a 4 credit course with MATH 220 Calculus I and MATH 221 Calculus II prerequisites; however, many students take the course after completing MATH 222 Vector Calculus. It is a requirement for most engineering track students, mathematics majors and a popular elective for students with a mathematics minor.

The data used in this study was taken from the fall 2020 semester. The course ran asynchronously with asynchronous, synchronous and in-person options for recitations and labs. Lectures were posted asynchronously on Monday and Wednesday. Recitations ran on Thursdays during a 50 minute class period lead by a graduate student with the purpose of re-enforcing lecture topics by actively practicing lecture topics in practice problems and homework problems. Labs ran on Tuesdays for another 50 minute class period lead by a graduate student with the purpose of putting a visual or real world application to the mathematical concepts being learned.

The course was organized online through Canvas using modules. Modules were organized by topic in the chronological order they appeared in lectures and included links for lecture videos, PDF's for lecture notes and assignments.

In addition to lecture videos, students were given two open-source textbook options which have some differences in topic coverage: Notes on Diffy Qs - Differential Equations

Assignment Type	Description
Written Homework	Due once a week with either an in-person, synchronous or asynchronous recitation
Online Homework	Due once a week with either an in-person, synchronous or asynchronous recitation
Exams	3 midterms and one final
Quizzes	Due within a few days of lecture videos being posted
Labs	Due once a week with either an in-person, synchronous or asynchronous lab.

Table 1.2: MATH 340 Graded Assignments

for Engineers by Jiri Lebl, Elementary Differential Equations by Andrew Bennett.

Notes on Diffy Qs	Elementary Differential Equations
1st Order Equations	1st Order Linear Equations
Higher Order Linear ODE's	Higher Order Linear Equations
Fourier Series and PDE's	Laplace Transforms
More on Eigenvalue Problems	Series Solutions
The Laplace Transform	
Power Series Methods	
Nonlinear Systems	

Table 1.3: MATH 340 Chapters by Textbook

Chapter 2

Methodology

2.1 Dataset Description

Data was collected automatically through the learning management system Canvas which included extensive data on student canvas activity.

Report Data	**Definition**
Total Views	The total number of times the presentation was watched
First Watched	Date the presentation was first viewed
Last Watched	Date the presentation was last viewed
IP address	IP address of the computer used to view the presentation
User	Username used to view the presentation
Live Views	The number of times the presentation was watched live
On-Demand Views	The number of times the presentation was watched on demand
Time Watched	The total amount of time (hh:mm:ss) the user spent watching the presentation
Peak Connections	The highest number of concurrent views for the presentation
Trends	An intensity graph that highlights which parts of the presentation were viewed most
Platforms	Graphical data showing the top web browsers and operating systems used to view the presentation
Coverage	The amount of the presentation watched by a user

Table 2.1: Mediasite data collection table

We focused on data around lecture videos, namely which videos were being watched, for how long, for how many times and when. Our main data set for each course was examining how much each student viewed each video in terms of percent watched as seen in Table 2.2.

	Video ID		
Student ID	V_1	$\cdots$	V_n
S_1	98%	$\cdots$	100%
S_2	100%	$\cdots$	61%
$\vdots$	$\vdots$		$\vdots$
S_m	16%	$\cdots$	0%

Table 2.2: Student Coverage Table Format

2.2 Mathematics Background

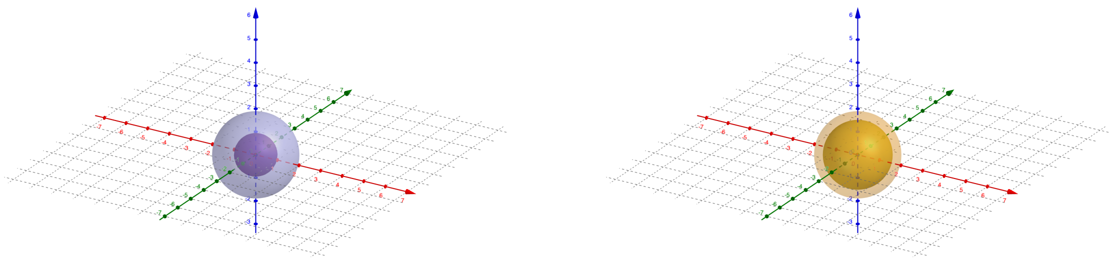
2.2.1 Dimension Reduction

Dimensionality causes issues for data analytics not only computationally, but visually. We cannot geometrically represent data in dimensions beyond 3, and it is easy to show that data analysis can change meaning as dimensions are reduced.

Take an Orange and a Plum. We might wonder about their ratio of edible portions to non-edible portions. An orange has a peel and a plum has a pit which reduce the overall ratio of edible portions on each. We can calculate this by hand easily by considering the volume of the fruit in ratio with the volume with the peel or pit. The question is what happens if we reduce the dimensional lens we use to view the fruit? In the 3D case the plum and orange are both a unit sphere, but have the following differences: The orange has a peel thickness of 0.15, the plum has a pit radius of 0.5. Our volumes work out to tell us that the plum has 87.5 percent edible parts while the orange has only 34.3 percent edible parts.

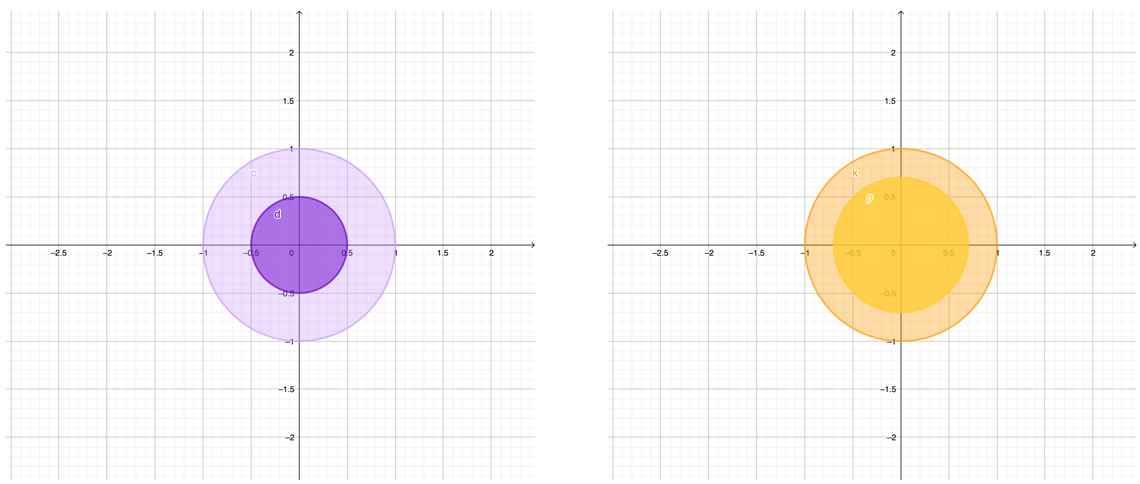
When we reduce this down to 2D we find we have different fruit to not-fruit ratios by area. In this case we have a plum that is 75 percent edible portion by area and for the orange we have 49 percent edible portion by area.

Now in the one dimensional line case of this problem we have a 50/50 ratio of edible to non-edible parts for the plum by distance and a 70/30 ratio of edible to non-edible parts



(a) Plum (b) Orange

Figure 2.1: 3D Plum and Orange

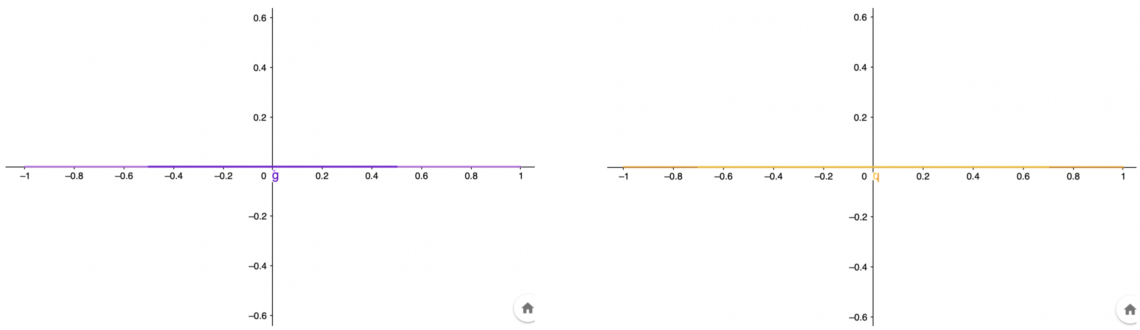


(a) Plum (b) Orange

Figure 2.2: 2D Plum and Orange

for the orange by distance. Note that these are small dimensional examples as we cannot geometrically visualize beyond 3 dimensions. This problem in a higher dimensional context will push data to the edges of the ball telling us that an orange is all peel and a plum is all fruit. This is obviously an issue, and especially for clustering when we are trying to pick out features.

It is easy to see now that information can tell us different things about characteristics depending on the dimensional lens we view it through. This phenomenon was originally described by Richard Bellman as " The Curse of Dimensionality". That is the phenomenon where things happening in high dimensions are not always happening in lower dimensions.[14]



(a) Plum (b) Orange

Figure 2.3: 1D Plum and Orange

Utilizing dimension reduction we can begin to recast our view of educational data into a space that can hopefully tell us something new about our data characteristics. Our space for MATH 570 is $\mathbb{R}^{36} \times \mathbb{R}^{175}$, and our space for MATH 340 is $\mathbb{R}^{206} \times \mathbb{R}^{129}$. Both of these are high dimensional spaces that are good candidates for dimension reduction. We will do this using principal component analysis.

Principal components are a collection of points in a real coordinate space that are a sequence of p unit vectors, where the i-th vector is the direction of a line that best fits the data while being orthogonal to the first i-1 vectors. Principal Component Analysis (PCA) is a multivariate technique in which a number of related variables are transformed into a smaller set of uncorrelated variables.[15] The following are the steps in computing the principal components of a given matrix.

The first step is to standardize the data. This allows the data to be manipulated on the same scale. We do this by computing a z-score.

$$z_i = \frac{x_{i,j} - \bar{x}_{ij}}{\sigma_{i,j}}$$

where x_i is the value for the i-th data point in the j-th column, $\bar{x}_i$ is the mean value for the j-th column and σ_{ij} is the standard deviation for the j-th column.

The second step is to compute the covariance matrix. Covariance provides a measure of the strength of the correlation between two or more sets of random variables. The covariance for two random variates X and Y, with sample size N and respective means $\bar{x}$ and $\bar{y}$ can be written as

$$cov(X,Y) = \sum_{i=1}^{N} \frac{(x_i - \bar{x})(y_i - \bar{y})}{N}$$

This step gives us a look at the relationship between the students data. This matrix is a $r \times r$ symmetric matrix where r is the number of dimensions and the entries are covarances associated with all possible pairs of initial variables.

$$S = \begin{bmatrix} s_1^2 & s_{12} & \cdots & s_{1r} \\ s_{12} & s_2^2 & \cdots & s_{2p} \\ \vdots & \vdots & & \vdots \\ s_{1r} & s_{2r} & \cdots & s_r^2 \end{bmatrix}$$

where s_i^2 is the variance and the covariance is

$$s_{ij} = \frac{n \sum x_{ik} x_{jk} - \sum x_{ik} \sum x_{jk}}{[n(n-1)]}$$

where the index k goes over the entire sample n.

The third step is to compute the eigenvectors and eigenvalues of the covariance matrix S. Since S is a $r \times r$ matrix we will have r eigenvalues $\lambda_1, \lambda_2, \cdots, \lambda_r$ which are obtained using the discriminant

$$|S - \lambda I| = 0.$$

The corresponding eigenvectors $v_1, v_2, \cdots, v_r$ are found by solving

$$(S - \lambda_i I)v_i = 0.$$

The fourth step is to order the eigenvectors by their eigenvalues in descending order. We then need to decide how many vectors to keep as the lower the eigenvalues the lower the significance. Our feature vector will be the eigenvectors we decide to keep.

The last step is to recast the data along the principal components axes by multiplying the transpose of the feature vector of our chosen eigenvectors by the transpose of our standardized original data set.[15]

This process is particularly taxing computationally. This complexity can be measured by something we call big-O notation which gives an upper-bound on a function. Consider a function $f(n)$ that is non-negative for all $n \geq 0$. We say that $f(n)$ is "big-O" of $g(n)$ if for every $n_0 \geq 0$ and constant $c > 0$ such that $f(n) < cg(n)$ for every $n \geq n_0$. We can write this in set notation as

$$O(g(n)) = \{f(n) | \exists c > 0 \text{ and } n_0 \text{ s.t. } \forall n \geq n_0, 0 \leq f(n) \leq cg(n)\}$$

In measuring the computational complexity of PCA we will consider it in two parts: Computing the covariance matrix and the eigenvalue decomposition. The covariance matrix computation is $O(p^2 n)$, and eigenvalue decomposition computation is $O(p^3)$. Together for PCA these come out to $O(p^2 n + p^3)$. We are able to speed-up this process by substituting in singular value decomposition in place of eigenvalue decomposition.

The singular value decomposition (SVD) of a matrix A with n rows and m columns is

$$A = U \Sigma V'$$

where U is a $n \times r$ matrix, Σ is a diagonal matrix with r singular values and V' a $r \times m$ matrix.

$$A = \begin{bmatrix} u_{11} & \cdots & u_{1r} \\ \vdots & & \vdots \\ u_{m1} & \cdots & u_{mr} \end{bmatrix} \begin{bmatrix} \sigma_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_r \end{bmatrix} \begin{bmatrix} v_{11} & \cdots & v_{1n} \\ \vdots & & \vdots \\ v_{r1} & \cdots & v_{rn} \end{bmatrix}$$

The SVD of a matrix is such a useful tool because it exists for all matrices, and it is unique so long as the singular values are unique. This gives SVD the additional reputation of being more accurate than other methods. The applications are wide, but for our purposes SVD serves mainly as a tool to decrease the computational complexity and increase the accuracy of finding principal components. This is important to note as the PCA function in the statistical program R uses SVD in its computation of principal components. The computational complexity for SVD is $O(\min\{pn^2, p^2n\})$ which is preferable over eigenvalue decomposition $O(p^3)$.

We are able to make this swap because the eigenvalues obtained through SVD are the same eigenvalues of the covariance matrix S obtained for PCA. They are related through the singular values σ_i of the diagonal matrix Σ in the SVD of our data matrix. SVD can be computed as follows:

Given an $n \times m$ matrix A, to find the SVD you first need to find the transpose of A denoted A^T. You then find AA^T and compute the characteristic polynomial $det(AA^T - \lambda I)$ to find the singular values $\sigma_1, \sigma_2, \cdots, \sigma_i$ for $i = 1, 2, \cdots, n$. You then compute the right singular vectors v_i by plugging in each of the singular values and row reducing to unit-length vectors. To compute the left singular vectors u_i you can use the equation $u_i = \frac{1}{\sigma} A v_i$.

This method, while functional for computing by hand, is taxing for large matrices.

R, using the LAPACK package, uses the following algorithm if the matrix is square:

1. Reduce the matrix A to bi-diagonal form $A = U_1 B V_1^*$ where U_1 and V_1 are unitary matrices.

2. Transform B to diagonal form Σ using two sequences of unitary matrices: $B = U^k \Sigma [V^k]^*$.

3. Combine the first two steps to approximate A as $A = U_1 U^k \Sigma [V^k]^* V_1^*$ where $U = U_1 U^k$ and $V^* = [V^k]^* V_1^*$.

If the matrix is not square with more rows than columns, it does QR factorization and then performs the above SVD algorithm on R. The SVD of A is then $A = (QU)\Sigma V^*$. If the matrix is not square with more columns than rows, it does LQ factorization and perfoms the SVD algorithm on L. The SVD of A is then $A = U\Sigma(V^*Q)$.[16]

2.2.2 Clustering

Clustering algorithms allow us to analyze our data by comparing data points and grouping them by similarity or closeness. One of the most basic clustering algorithms is K-means clustering. This is an iterative 2-step algorithm:

1. To obtain k clusters, arbitrarily choose k centroids and assign each data point to a cluster.

2. Compute the new centroids of each cluster.

This is repeated until clusters are optimized. This happens when the clusters are no longer changing with each iteration.

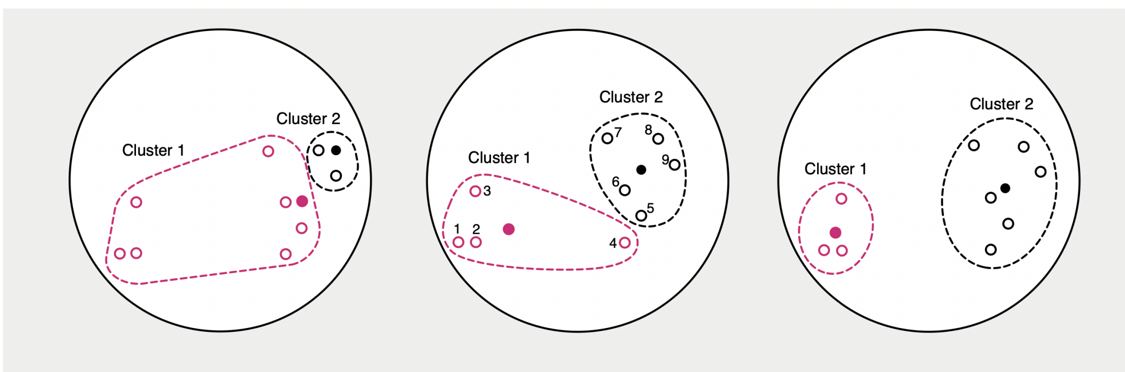


Figure 2.4: Standard K-Means clustering visualization by Vance Faber[1]

The trouble with K-means clustering is that the use of computed centrioids can be vulnerable to extreme data. A similar, but more robust algorithm is partitioning around medoids (PAM).

To obtain k clusters, k representative objects are selected and the remaining objects are assigned to a cluster based on their closeness to the representative object. Therefore a tight cluster tells us there are strong similarities in this cluster while a loose cluster will tell you that there are weaker similarities. These distance measures are minimized to give the "tightest" set of clusters possible. This closeness distance is measured using Euclidean distance.

$$d(i,j) = \sqrt{(x_{i1} - x_{j1})^2 + (x_{i2} - x_{j2})^2 + \cdots + (x_{ir} - x_{jr})^2}$$

This is an algorithm in two phases: BUILD and SWAP. In the BUILD phase we are finding the k representative objects in the data.[17]

1. Consider an object i which has not yet been selected
2. Consider a non-selected object j and calculate the difference between its dissimilarity D_j with he most similar previously selected object and its dissimilarity $d(i, j)$ with object i
3. If this difference is positive, object j will contribute to the decision to select object i. Therefore we calculate $C_{i,j} = \max(D_j - d(j, i), 0)$
4. Calculate the total gain obtained by selecting object i: $\sum_j C_{ji}$
5. Choose the not yet selected object i which maximizes $\max_i \sum_j C_{ji}$

Once k objects have been found we begin the SWAP phase.

1. Consider a non-selected object j and calculate its contribution C_{jih} to the swap:
   (a) If j is more distant from both i and h than from one of the other representative objects, C_{jih} is zero
   (b) If j is not further from i than from any other selected representative object ($d(j, i) = D_j$), then two situations need to be considered:
      i. j is closer to h than to the second closest representative object $d(j, h) < E_j$ where E_j is the dissimilarity between j and the second most similar representative object. In this case the contribution of the object j is the swap between objects i and h is $C_{jih} = d(j, h) - d(j, i)$
      ii. j is at least as distance from h than from the second closest representative object $d(j, h) \geq E_j$. In this case the contribution of the object j to the swap is $C_{jih} = E_j - D_j$
   (c) j is more distant from object i than from at least one of the other representative objects but closer to h than to any representative object. In this case the contribution of j to the swap is $C_{jih} - d(j, h) - D_j$

2. Calculate the total result of a swap by adding the contributions $C_{jih} : T_{ih} = \sum_j C_{jih}$

3. Select the pair (i, h) which minimizes $\min_{i,h} T_{ih}$

4. If the minimum T_{ih} is negative, the swap is carried out and the algorithm returns to step 1. If the minimum is positive or 0, the value of the objective cannot be decreased by carrying out a swap and the algorithm stops.

This function returns a representative for each cluster along with their principal component values in addition to a cluster vector which gives the cluster assignment for each data point. From this we can start to parse what behaviors these components represent and how these cluster together to give us possible student behavioral profiles.

For both PAM and K-means clustering we need to tell the algorithm how many clusters we want. This step is crucial, and I chose to solve this problem using hierarchical clustering.

Hierarchical clustering using the hclust function in R uses a bottom-up, agglomerative, method of clustering. All objects start as their own cluster being distinct from every other object. The program then joins these clusters by similarity, iteratively, until all objects are assigned to one encompassing cluster. The complete-linkage method is used along with Euclidean distance.

Complete-linkage method uses the farthest distance between members of a cluster to calculate distance. That is, mathematically we have $D(X, Y) = \max_{x \in X, y \in Y} d(x, y)$ where X and Y are the two clusters, x and y are the members of each cluster and the distance $d(x, y)$ is the Euclidean distance. All distances between members are calculated and the maximum distance is used.

The hclust algorithm is as follows: Starting with a distance matrix D that has distances $D(i, j)$.

1. The first cluster level is represented by $L(0) = 0$ and the sequence number is $m = 0$

2. Out of the current clusters, find a cluster pair X and Y according to $d(X, Y) = \min D(i, j)$

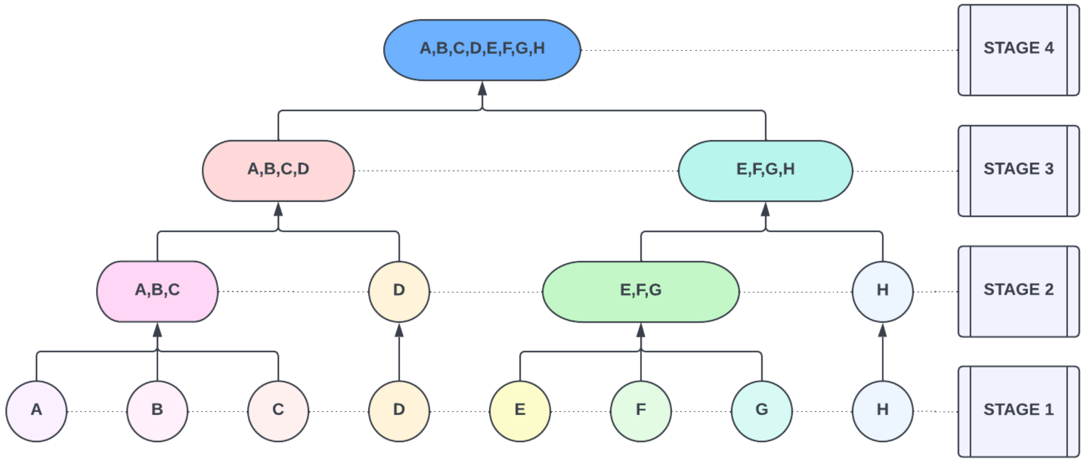


Figure 2.5: Hierarchical clustering process visualized as a Dendrogram

3. Move onto the next sequence number m=m+1 and merge clusters X and Y into a single cluster to form the next set of clusters.

4. Update the distance matrix D with the distances between the new clusters. The distance between the new cluster k and the old is $d(X, Y) = \max d(k, X), d(k, Y)$

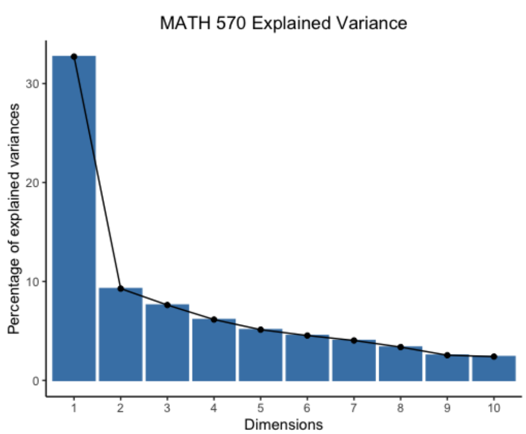
5. If all objects are in one cluster, STOP. Else, repeat steps 2-5.

This method can be helpful for getting an initial idea on how many clusters are appropriate for the data. Graphing this data gives you a visual aid in seeing the cluster options at different levels. The more isolate a cluster looks, the more likely it will be a good cluster for further analysis. I use this as a basis for how many clusters I should use when moving onto the next step in using PAM.

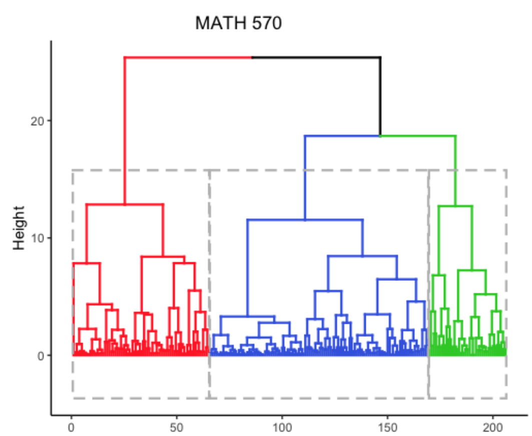
Chapter 3

Discussion and Further Research

3.1 Results: MATH 570



(a) MATH 570 PCA Explained Variance

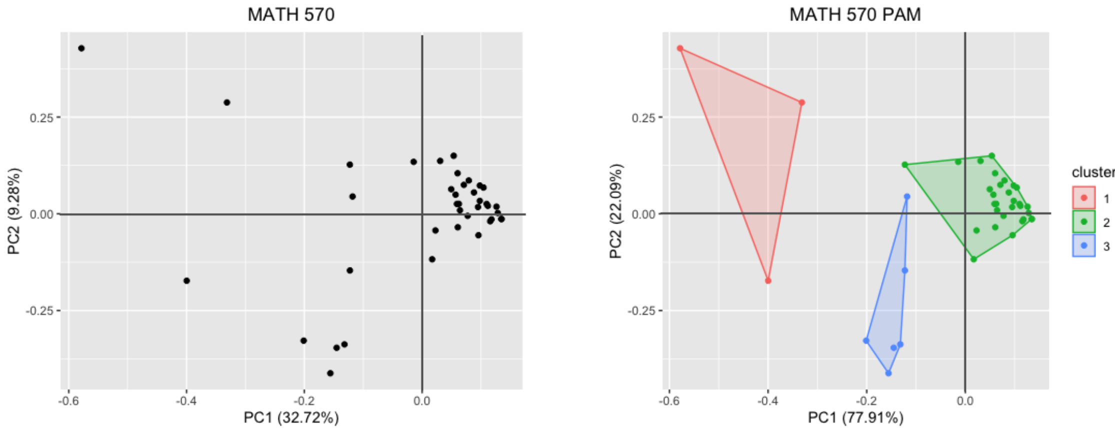


(b) MATH 570 dendrogram

Figure 3.1: MATH 570 Pre-processing

The PCA using the prcomp() function showed that the explained variance was largely projected onto the first two principal components as seen in figure 3.1(a). Using hierarchical clustering, the program chose 3 clusters as the optimal choice of k for PAM as seen in figure 3.1(b). In the figure, the grey segmented boxes map out the optimal clusters.

The principal components represent different features of student video usage. Principal component 1 (PC1) represents how much students watched posted videos with a negative



(a) before PAM

(b) after PAM

Figure 3.2: MATH 570 Clustering Results

value indicating lots of video usage and a positive value indicating less video usage.

Principal component 2 (PC2) represents that type of video a student watched. This comes down to video weights. Some videos are given negative weight, and students who watched a lot of those videos will appear on the negative end of PC2. Other videos have positive weights, and students who watched a lot of those videos will appear on the positive end of PC2. We define videos with a positive weight as Type A videos and videos with a negative weight as Type B videos. To better understand the differences in these videos, we have the following table matching the titles of the 10 most extreme positive weight and negative weight videos.

The visual distinction on PC2 is made even more clear if we re-run our PCA and PAM on non-normalized data. This changes the distance measure between points. What we end up with is a version that is scaled out. Note that PCA assigns sign arbitrarily, so in this re-run of the data the videos weights switch sign.

With this new view we get three distinct groups. Cluster 1 is students who watched a lot of videos. Cluster 2 Students who did not watch a lot but what they did watch was more B type videos. Cluster 3 is Students who did not watch a lot but what they did watch was more A type videos.

Breaking these into outcome categories, high performing students either watched a lot

Positive Most Videos		Negative Most Videos	
Video Title	Video Wt	Video Title	Video Wt
HM Day 26 - Afterward	0.09962913	Practice Final - Prob 4	-0.242934
HM Day 30 - Cubic	0.09301993	Practice Final - Prob 3	-0.2402363
HM Day 30 - Depression	0.09157531	Practice Final - Prob 9	-0.2378444
HM Day 28 - Lattice Multiplication	0.08519801	Practice Final - Prob 1	-0.2363793
HM Day 40 - Leibniz	0.07668704	Practice Final - Prob 6	-0.2362133
HM Day 11 - Eratosthenes	0.07245274	Practice Final - Prob 8	-0.2355863
HM Day 42 - HW	0.07222977	Practice Final - Prob 5	-0.2342447
HM Day 10 - History	0.07212177	Practice Final - Prob 7	-0.2205568
HM Day 28 - Discussion	0.07165746	Practice Final - Prob 2	-0.2062629
HM Day 28 - Excess	0.06976004	HM Day 14 - Prob 5	-0.1709901

Figure 3.3: Top 10 Type A and Type B Videos

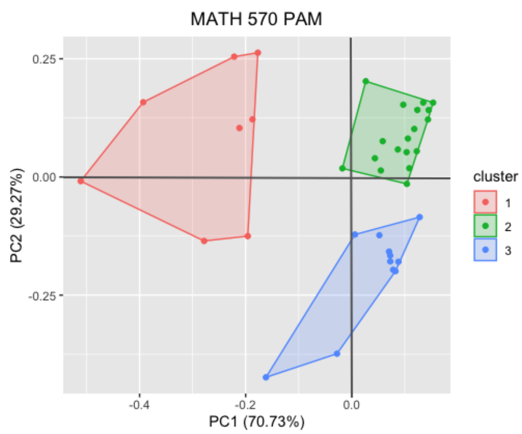
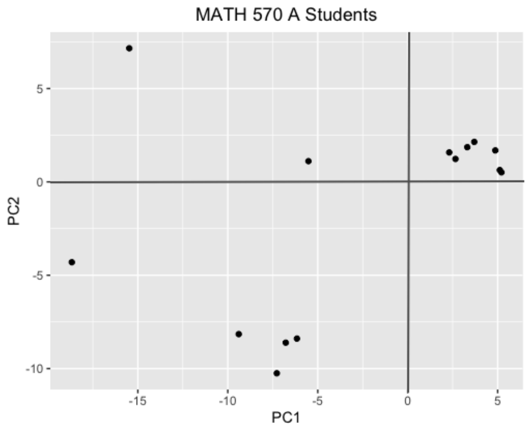
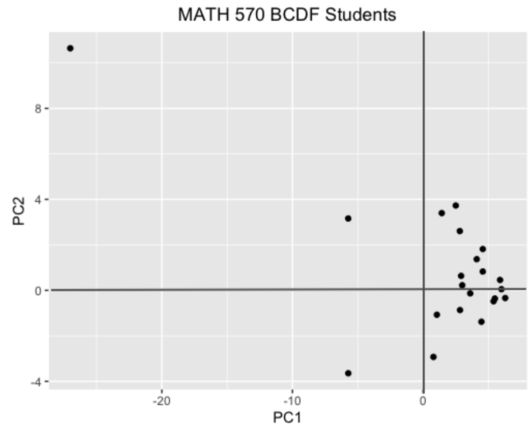


Figure 3.4: PAM on non-normalized data

of videos or did not watch many videos but focused on Type A videos. Lesser performing students, in general, did not watch many videos, but the videos they did watch were either Type A or Type B. Most of the students watching type A in this category were B students while most of the students watching B videos were students more significantly struggling in the course.



(a) A students

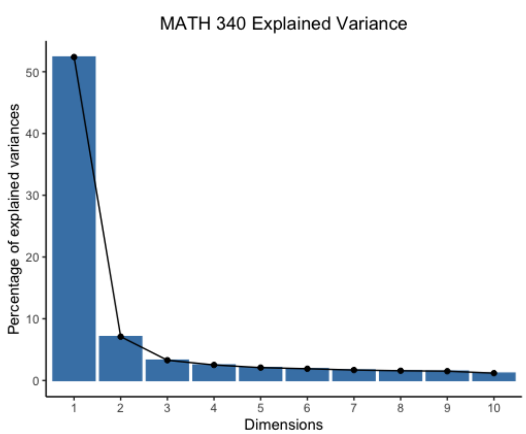


(b) Other Students (BCDF)

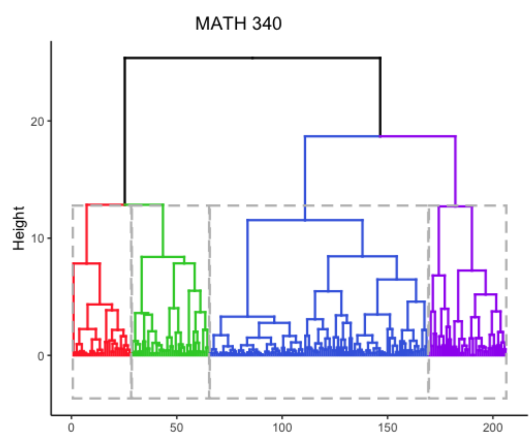
Figure 3.5: MATH 570 Results by Grade Outcome

3.2 Results: MATH 340

The PCA showed that the explained variance was overwhelmingly projected onto the first component as seen in figure 3.6(a) with less than 10 percent being projected onto the second component. Using hierarchical clustering, the program chose 4 clusters as the optimal choice of k for PAM as seen in figure 3.6(b).



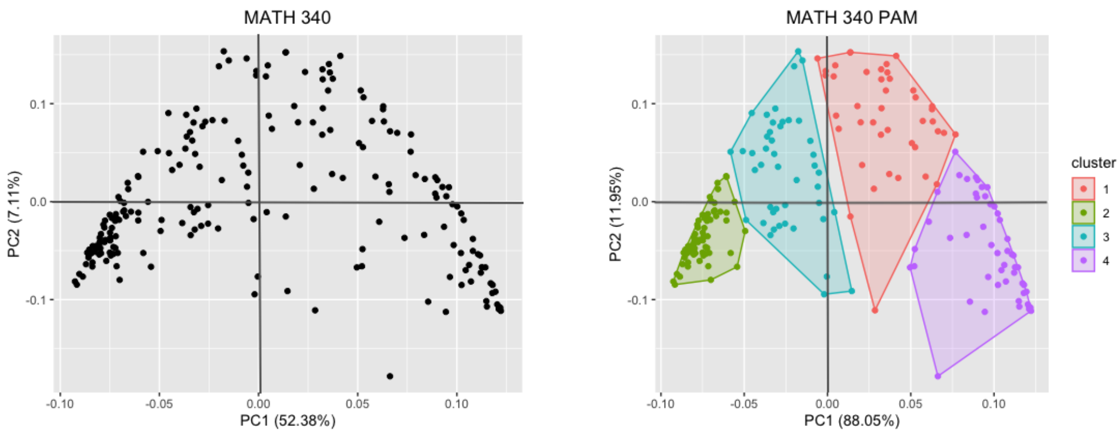
(a) MATH 340 PCA Explained Variance



(b) MATH 340 Dendrogram

Figure 3.6: MATH 340 Pre-PAM

Plotting the data using the first two principal components, the data takes on a parabolic shape as seen in figure 3.7(a) with a best fit line of a flat horizontal line. This indicates components that are heavily dependent, which is likely because of the strong variance on



(a) before PAM (b) after PAM

Figure 3.7: MATH 340 Clustering Results

PC1.

Based on analysis of the student watch data, PC1 measures the amount a student watched over the semester. Students with a negative PC1 value watched more over the semester while students with a positive value watched less. Clusters 2 and 3 both have students who watched more over the semester whereas clusters 1 and 4 have more students who watched less over the course of the semester.

PC2 measures when in the semester they watched the most videos. A negative value on PC2 shows students who are watching a more consistent amount either consistently a lot or consistently very little. A positive PC2 value shows students who were watching more in the beginning of the semester and then drop off toward the end.

Plotting the views per video over the course of the semester gives us further insight into these results. Consider Figure 3.7 above, the trend over the course of the semester is a steady decline in views. This trend could be strong enough to shape the data like we saw in Figure 3.6.

Looking at video reviews rather than views we can re-sort our data for A students and other boarder-line students. Consider the data shown in Figure 3.8. This shows the top 10 most reviewed videos for A students next to the to 10 most reviewed videos for B students.

These have a lot of overlap, so going further we look at if each of these videos were

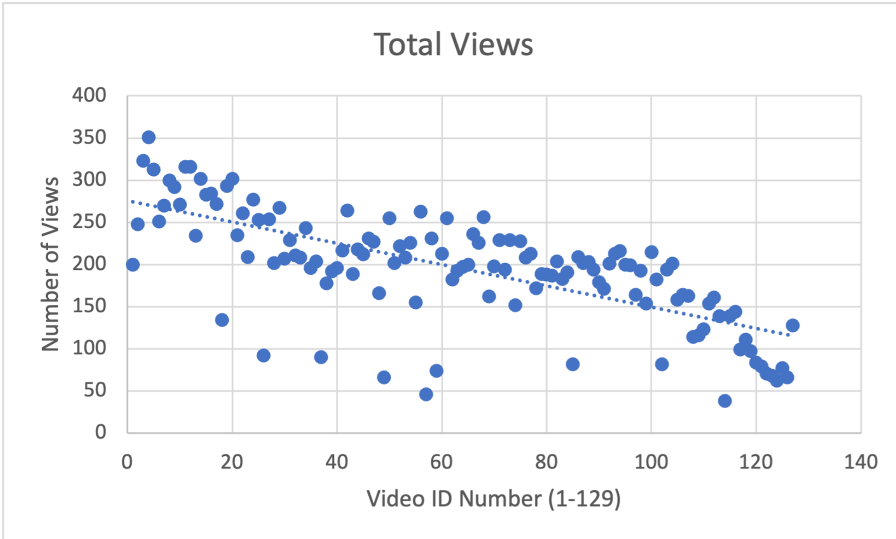


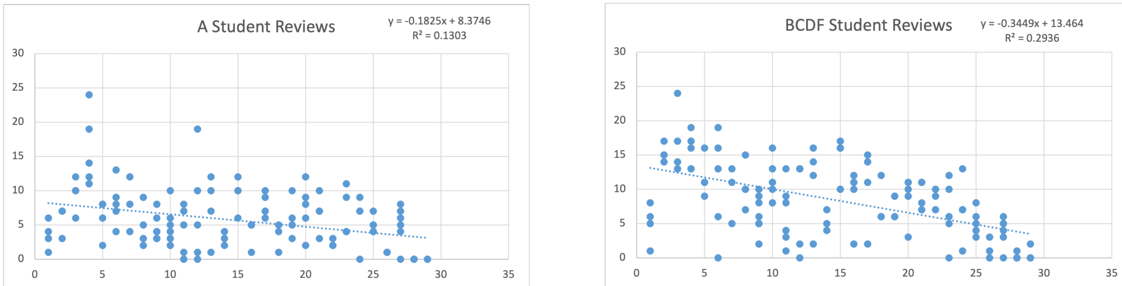
Figure 3.8: MATH 340 Total Views Per Video

A Student Reviews		Other Student Reviews	
Vid ID	Vid Title	Vid ID	Vid Title
12	DE Lec 3 - Paradigm	5	DE Lec 2 - First Example
13	DE Lec 4 - Bernoulli Paradigm	11	DE Lec 3 - Introduction
14	DE Lec 4 - Homogeneous Equations	12	DE Lec 3 - Paradigm
15	DE Lec 4 - Homogeneous Paradigm	13	DE Lec 4 - Bernoulli Paradigm
22	DE Lec 6 - Paradigm	14	DE Lec 4 - Homogeneous Equations
27	DE Lec 7 - Picard Iteration	18	DE Lec 5 - Pop Models
54	De Lec 12 - Homework	20	DE Lec 6 - Exact Equations
59	DE Lec 13 - Undamped Forced	22	DE Lec 6 - Paradigm
66	DE Lec 15 - Paradigm	66	DE Lec 15 - Paradigm
125	DE Lecture 4 - Bernoulli Equations	125	DE Lecture 4 - Bernoulli Equations

Figure 3.9: General Top Reviews by Student Outcome Group

watched more by one group or the other. To do this we break down the class into two groups: A students and BCDF students. With each of these, we take their review data for each video and create a linear model predicting how much the students are expected to review a particular video. This model comes out as seen in Figure 3.9 below.

Using these linear models we can find z scores for each of the videos for each student group by finding $z = \frac{x-e}{s}$ where x is our actual data point for how much the video was reviewed by students in the group, e is our expected number of reviews based off the linear model and s is the standard deviation. To combine these z-scores together, we can then



(a) A video reviews (b) BCDF video reviews

Figure 3.10: Expected Video Reviews

find the differential by finding the difference of the z-score for each video in the A student group and subtracting the z-score for each video in the BCDF student group. This gives us one weighted value for each video which describes whether A student or BCDF students watched more with positive weight telling us A students watched more, negative weight telling us BCDF watched more and values around zero telling us they watched about the same amount.

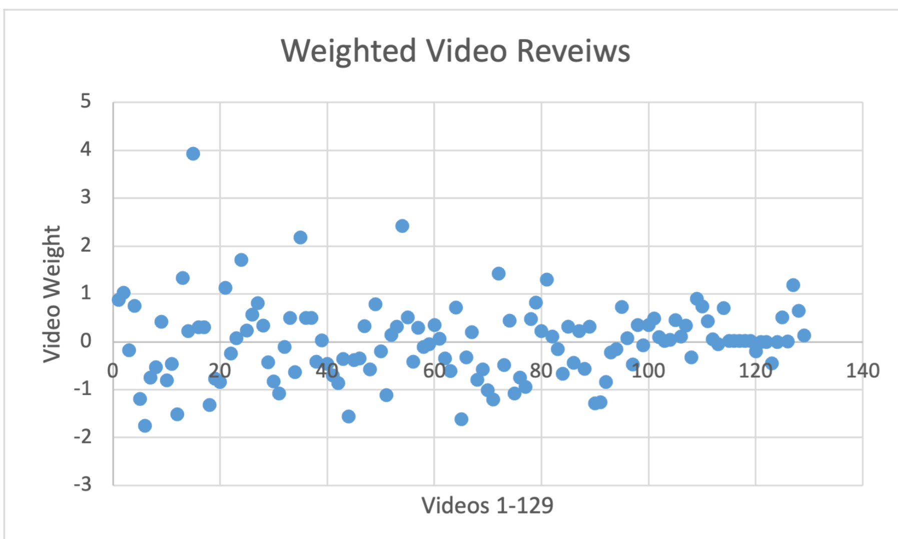


Figure 3.11: Weighted Video Reviews

Plotting these we can see a conic shape. The videos are more polarized between the student groups toward the beginning of the semester and start to even out toward a weight of zero as the semester progresses. Results of this are shown in figure 3.11.

Most reviewed by A students		Most reveiwed by BCDF students	
Vid ID	Vid Title	Vid ID	Vid Title
15	DE Lec 4 - Homogeneous Paradigm	6	DE Lec 2 - Initial Value Problem
54	De Lec 12 - Homework	65	DE Lec 15 - Linearity
35	DE Lec 9 - Extra Credit 2	44	DE Lec 10 - Sines and Cosines
24	DE Lec 6 - XC 1	12	DE Lec 3 - Paradigm
72	DE Lec 17 - Conserved Quantities	18	DE Lec 5 - Pop Models
13	DE Lec 4 - Bernoulli Paradigm	90	DE Lec 22 - BVP
81	DE Lec 20 - Example 6-2-2	91	DE Lec 22 - Heat Eqn
127	DE XC 5	71	DE Lec 17 - Classifying Critical Points
21	DE Lec 6 - IVP	5	DE Lec 2 - First Example
2	DE Lec 1 Class Procedures	51	DE Lec 12 - Building the Equation

Figure 3.12: Top 10 Most Reviewed by Student Outcome Group

3.3 Discussion

3.3.1 Question 1

Can we identify patterns in student behavior by their online utilization of lecture videos?

The online behavior for students in MATH 570 History of Math was their viewing habits of posed recordings of synchronous lecture videos. These views were either reviews or initial viewings for students who did not attend the synchronous lecture. The patterns found were: How much students watched and what type of video, A Type or B Type students watched.

For MATH 340 Elementary Differential Equations the viewing habits of posted lecture videos were: How much students watched and when they watched. It seems the strong linear decline in video watching over the course of the semester was such a strong factor that it overpowered all other features from the data. This could be accounted for by the overall homogeneity of the class which is overwhelmingly engineering students. These students have the same level of background and reasonably might have very similar behaviors in their math classes.

3.3.2 Question 2

Are there specific behavioral patterns from students who succeed vs. do not succeed in the course?

Defining features for students in MATH 570 were how much they watched and what they were watching. High achieving students were generally working really hard reviewing a lot of the videos or reviewing less, but picking specific types of videos. We called this video type A type videos, which were videos focused on explaining topics. Borderline and struggling students tended not to review a lot of videos, and the videos they did review were B type videos. These videos are ones explaining how to do particular problems. These patterns suggest that students who are only reviewing videos to complete homework are less likely to succeed than students who are reviewing videos to learn a concept.

There were no defining features for MATH 340 students beside some students working really hard watching all the videos did well. However, there were also plenty of successful students who never watched lecture videos. The lecture videos, in general, were not required for the course. Two separate textbooks were also made available, so students who preferred to work from a book could do so without needing the lectures. Students, of course, also had access to the web in which topics in Elementary Differential Equations are easy to find.

This lead us toward a new question: viewing this through the same lens we looked at MATH 570, how are MATH 340 students *reviewing* videos?

Like the general watching trend, we had students who reviewed a lot of videos and some that reviewed very few. In addition, as we would expect, reviews also decreased over the course of the semester. However, looking separately at reviews for A students versus Other students we were able to find a linear model for predicting how much we would expect an A student or Other students to review a particular video. We used this model to calculate z-values for each video and then the differential between A students and Other students.

Comparing the videos most watched by A students versus Other students it became apparent that struggling students are reviewing videos on the basics within the first few weeks the most whereas students doing well are watching videos on more complex topics or optional videos like extra credit.

3.3.3 Future Research

The data for this thesis is limited to students at Kansas State University in two specific mathematics courses. Building on these findings, mining data from other student populations in mathematics courses could show other interesting results. Since online learning hit universities around the world at about the same time, there is an abundance of potential educational data.

Reproduction of this thesis will prove difficult as the conditions of this data are unique to the COVID-19 situation. We have never, and might not ever again, see this many students learning online. However, online learning will not be going anywhere. These flexible modalities are still a popular option, and more can be learned even as the student population taking them change. Using these data mining techniques along with student survey participation could further insight into why students are making particular behavioral choices.

Another promising direction would be to focus on how we can use these data mining techniques to detect students struggling in an online setting early in the semester. The results from the Elementary Differential Equations class, in regard to video reviews, seem to point in this direction. The student perspective could be examined with access to student surveys and corrective student actions could be explored.

Bibliography

[1] Vance Faber. Clustering and the continuous k-means algorithm. *Los Alamos Science*, 22, 1994.

[2] Samuel Messick. *Cognitive Styles in Educational Practice*. Educational Testing Services, 1982.

[3] Roy D. Pea. Cognitive technologies for mathematics education. In A. H. Schoenfeld, editor, *Cognitive Science and Mathematics Education*, chapter 4, pages 89–122. Lawrence Erlbaum Associates, Inc, Hillsdale, NJ, 1987.

[4] Hassan Alamolhodaei. *A Study in Higher Education Calculus and Students' Learning Styles*. PhD thesis, Glasgow University, 1996.

[5] Kansas State University. Instructional mode and course modalities. Aug 2021. URL https://www.k-state.edu/registrar/instructionalmode/coursemodalities/index.html.

[6] Kansas State University. Course component type definitions. Aug 2021. URL https://www.k-state.edu/registrar/instructionalmode/types/index.html.

[7] Donatien N'Dri and Mario Lefebvre. Teaching differential equations online. *Atti della Accademia Peloritana dei Pericolanti - Classe di Scienze Fisiche, Matematiche e Naturali*, 99(S1):12, 2021. ISSN 1825-1242. doi: 10.1478/AAPP.99S1A12. URL https://cab.unime.it/journals/index.php/AAPP/article/view/AAPP.99S1A12.

[8] Meehyun Yoon, Jungeun Lee, and Il-Hyun Jo. Video learning analytics: Investigating behavioral patterns and learner clusters in video-based online learning. *The Internet and Higher Education*, 50:100806, 2021. ISSN 1096-7516. doi: https://doi.org/10.1016/j.

iheduc.2021.100806. URL https://www.sciencedirect.com/science/article/pii/S1096751621000154.

[9] OVAE; Tanya Shuy and TEAL staff. URL https://lincs.ed.gov/sites/default/files/3_TEAL_Self%20Reg%20Learning.pdf.

[10] Yun-Wei Chu, Elizabeth Tenorio, Laura Cruz, Kerrie Anna Douglas, Andrew S. Lan, and Christopher G. Brinton. Click-based student performance prediction: A clustering guided meta-learning approach. *2021 IEEE International Conference on Big Data (Big Data)*, pages 1389–1398, 2021.

[11] Gianni Fenu and Roberta Galici. Modelling student behavior in synchronous online learning during the covid-19 pandemic. *CEUR Workshop Proceedings: First International Workshop on Enabling Data-Driven Decisions from Leaning on the Web*, 2876, 2021.

[12] Kansas State University. Math 570 - history of mathematics - acalog acms™. Aug 2018. URL https://catalog.k-state.edu/preview_course_nopop.php?catoid=2&coid=5829.

[13] Kansas State University. Math 240 - elementary differential equations - acalog acms™. Aug 2018. URL https://catalog.k-state.edu/preview_course_nopop.php?catoid=1&coid=15846.

[14] Richard Bellman. *Adaptive Control Processes: A Guided Tour*. Princeton University Press, 1961.

[15] J. Edward Jackson. *A User's Guide to Principal Components*. Wiley Series in Probability and Mathematical Statistics, 1991.

[16] Susan Blackford. *LAPACK Users' Guide Third Edition*. SIAM, 1999.

[17] Leonard Kaufman and Peter J.Rousseeuw. *Finding Groups in Data*. Wiley Series in Probability and Mathematical Statistics, 1990.