Data mining and intervention in Calculus I

by

Ian Manly

B.S., University of Pittsburgh, 2013

M.S., Kansas State University, 2015

AN ABSTRACT OF A DISSERTATION

submitted in partial fulfillment of the
requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Mathematics
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2018

Abstract

Many students have difficulty performing well in Calculus 1. Since Calculus 1 is often the first math course that people take in college, these difficulties can set a precedent of failure for these students. Using tools from data mining and interviews with Precalculus and Calculus 1 students, this work seeks to identify the different types of students in Calculus 1, determine which students are at risk for failure, and to study how intervention can help them succeed both in mathematics and in their college careers.

Data mining and intervention in Calculus I

by

Ian Manly

B.S., University of Pittsburgh, 2013

M.S., Kansas State University, 2015

A DISSERTATION

submitted in partial fulfillment of the
requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Mathematics
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2018

Approved by:

Major Professor
Andrew Bennett

Copyright

Abstract

Many students have difficulty performing well in Calculus 1. Since Calculus 1 is often the first math course that people take in college, these difficulties can set a precedent of failure for these students. Using tools from data mining and interviews with Precalculus and Calculus 1 students, this work seeks to identify the different types of students in Calculus 1, determine which students are at risk for failure, and to study how intervention can help them succeed both in mathematics and in their college careers.

Table of Contents

List of Figures

List of Tables

Acknowledgments

I would like to thank Dr. Craig Spencer, Dr. Nathan Albin, Dr. Sherri Martinie, and Dr. Juan Du for being on my committee despite having numerous other responsibilities. Thanks also go to Kathy Brown, Reta McDermott, and Tina Orwig for helping me jump through the necessary hoops during my time at graduate school. Finally, many thanks go to my major advisor Dr. Andrew Bennett for providing me with extremely valuable research and career advice. I wouldn't have been able to navigate the job market and make it through the process of getting a PhD without his support.

Finally I would like to thank my parents John and Julia Manly, and my friends Kyle, Cari, Abby, Anna, and many others for being there for me while I was working on this dissertation - you guys are the best. I deeply appreciate having such wonderful people in my life.

Dedication

There have been many points in my life in which I wanted to give up on mathematics and do something else instead, yet here I am. I dedicate this thesis to all of the teachers, friends, and family members who have encouraged me to continue studying math, even when it was overwhelming difficult to do so.

Chapter 1

Introduction

In the past, when it was difficult to make copies of books or notes, professors and teachers used lectures to give their students the opportunity to make copies of the information in class Bligh (1998). This practice remains common in universities today, despite the multitude of textbooks and online resources containing the material. In particular, lower level university courses such as Calculus 1 can have hundreds of students per class. Given the sheer size of these classes, it's impossible for an instructor to get to know their students as individuals, and as such it's difficult to adjust their teaching style to accommodate the varying types of students in their course.

This situation can be particularly vexing for students in Calculus 1. Most of the students are freshman, and Calculus 1 is their first college level mathematics course. It can be overwhelming to transition from smaller, more intimate high school classes to massive, impersonal college lectures. The purpose of this study is to combat this phenomenon - how can we cater to the different types of students in Calculus 1, and in particular, how can we identify and assist those students who are struggling?

To this end, the author will use data mining techniques - in particular, clustering algorithms - to partition Calculus 1 based on assignment scores, their performance on the first midterm exam, and attendance. This needs to be done during the first few weeks of the semester as students become "set in their ways" and are less likely to change negative behaviors Cusino (2013). After calculating how many groups the course will be broken into, and the composition of these groups, the students will be interviewed to determine the personality traits and attitudes towards math that typical members of each cluster exhibit. Once it is known what types of students are in the course, it may be possible to differentiate instruction to more effectively cater to the individual needs of the students.

After performing this analysis, it may be the case that one of the clusters may contain a disproportionately large number of at-risk students. In an effort to study and assist these struggling students, the author will analyze a special version of Precalculus offered at Kansas State University called (internally, at least) “Bailout Precalculus”. If a student in (the fall semester of) Calculus 1 fails or does poorly on their first exam, they are offered the opportunity to switch into a newly created section of Precalculus, and have their scores and any official record of enrollment in Calculus 1 erased. They would then spend the remainder of the semester improving their algebra and trigonometry skills, as well as relearning the first few weeks of Calculus 1. The researcher believes that this course has a strong impact on the students enrolled in it, and wishes to find out what makes this class effective. This is the focus of the latter part of this study.

Motivation

What is to be gained by investigating the students of Calculus 1 and Precalculus? Calculus 1 certainly isn’t the only course with struggling students. Why did the author choose to study these courses in particular for this study?

For the vast majority of those enrolled in Calculus 1 (especially during the fall semester), it is their first experience of a college mathematics course. The other courses at Kansas State which are likely to be someone’s first college math course are College Algebra and Studio College Algebra (College Algebra with a computer lab component). Given the high enrollment for these courses, Dr. Rachel Manspeaker decided to use data mining in an attempt to classify the students and differentiate instruction Manspeaker (2011). She was successful, and found 5 different classifications of students, and offered extra assistance to those in one of the groups that contained many struggling students. In large part, the success of that study motivated the present one - can Calculus 1 also be partitioned into meaningful groups?

Unlike College Algebra, where it can be safely assumed that the students have seen a majority of the course material already in high school, it is not the case that all Calculus 1 students have had previous exposure to calculus concepts. However, for PhD granting institutions, 67% of students have taken calculus in high school Bressoud et al. (2015). This can be intimidating for the remaining 33% of students. While their peers are breezing through the early parts of the course, students who haven't taken calculus before may find themselves feeling not only lost and confused, but demoralized and inferior since they seem to be the only ones not understanding what's going on. The Bailout Precalculus course provides a remedy for this, since the final five weeks of the course are spent reviewing the first few chapters and sections of Calculus 1. This is one of the reasons why understanding this course is important - it levels the playing field between those students who have had previous calculus training and those who haven't.

This issue becomes especially relevant when it comes to advising incoming engineering students. For many of the departments in the School of Engineering at Kansas State, taking Calculus 1 in the first semester of freshman year is essential to completing the requirements for the degree in four years. This is due to the fact that Calculus 1 (along with the other courses in the Calculus sequence) is a prerequisite for many required engineering and science courses, such as Engineering Physics 1. As such, many advisers push students who aren't mathematically prepared to take Calculus 1 to keep on schedule. Also, many scholarships that students earn stop assisting with tuition after four years, adding additional pressure to prematurely enroll in Calculus 1.

First generation students are another group that may be adversely affected by the large lecture/recitation format of Calculus 1. Many first generation students have no idea what to expect when enrolling in a university. It is their first time away from home as adults, and they also have the added difficulty of not having anyone in their immediate family they can ask for guidance about college life. As such, large lecture courses can be overwhelming, especially in contrast to the smaller classes they may be used to from high school. This is another group that benefits from the Bailout Precalculus course. Rather than being thrown into a foreign situation and failing, first generation students can succeed in a course that is somewhere in between high school and college format, and be more prepared for the classes they will be taking in the future.

Finally, there is a monetary incentive to boosting retention through Bailout Precalculus. Every student that is retained in the university is a student who is continuing to pay tuition in one form or another. For every 1% increase in retention, the university saves $250,000 OPA (2011). Maintaining the effectiveness of this course is thus important in the modern era of higher education, when massive budget cuts are being implemented across the board.

In recent years, data mining techniques have become more common, both in research and in industry. In particular, clustering algorithms are useful for splitting large populations with many different traits into smaller groups, where the members of each group are similar to one another in some way. For example, in marketing, clustering algorithms are used to group consumers by buying preferences and demographics (An intro to clustering). Online video websites such as YouTube and Netflix use clustering algorithms to decide which videos or shows will appeal to which people. These algorithms are being used to do some impressive, cutting edge research, such as the development of voice recognition software(voice recognition site.)

In Rachel Manspeaker's thesis, she uses clustering algorithms (as well as interviews with representatives from each cluster) to divide the students of Kansas State's College Algebra course into five different clusters. This process was repeated over several different semesters to determine whether there were significant differences in the groupings from year to year or between the fall and spring semesters Manspeaker (2011). Despite these changes in the student population, there were no significant changes in the sizes and traits of the clusters from semester to semester. In light of this success, the researcher believes that such an approach will prove successful in another large, lower level math course at Kansas State.

However, why is it necessary to use data mining to differentiate instruction? Even with a large course, the lecturer and recitation instructors could tailor their lessons to accommodate different learning styles. While this approach may seem natural, there is little evidence to suggest that teaching for specific styles is effective, or evidence that students can accurately determine their own optimal learning style Vasquez (2009). Therefore, some other method (like data mining) must be used to differentiate instruction.

When attempting to split a population into groups, one has two options - a "white box" method or a "black box" method. A white box method uses expert knowledge and pre-exisiting frameworks to organize the groups (for example, grouping students based on learning styles), whereas a black box method solely relies on the data and properties of the population to determine what the makeup of the clusters should be (like using clustering algorithms from data mining) Manspeaker (2011). For this study, it is the opinion of the researcher that using black box methods will yield the most accurate and natural partition of the Calculus 1 population. The primary goal of clustering is differentiated instruction, and it would be best if membership in the clusters was based on the attributes of the students in the specific course.

After the clusters have been formed, the "personality" of each group will be determined by interviewing multiple students from each group. This, of course, is only possible provided that the students in each group do indeed have similar attitudes about the course, school, and mathematics as a whole. Since the clustering algorithms will only have as inputs the students' scores on assignments and exams, it may happen that the students in a cluster (or at least the ones interviewed) have little in common with one another. If that is the case, then it would be difficult to differentiate instruction based on that cluster. Even if the interviews yielded a consistent theme for the students in a group, it may still be difficult to positively impact that group in particular. For example, in Manspeaker's work, one group that emerged was dubbed "The Overachievers " for their high scores and strong work ethic Manspeaker (2011). However, despite having a consistent identity, there wasn't anything to be done for these students, as they were already succeeding in the course on their own! Therefore, depending on their particular attributes, it is possible that some (or perhaps all) of the clusters in this study may not have a consistent identity or be a suitable target for differentiated instruction.

In regard to Bailout Precalculus, there are several reasons why the researcher expects this course to positively influence retention and students' mathematical proficiency. First, since any grades or record of a student's enrollment in Calculus 1 is erased upon switching into Precalculus, a struggling student can begin their collegiate mathematical journey again without worrying about the impact of their prior performance on their grades. This can set the stage for a more optimistic view of both their college career in general and their progress in their STEM major specifically. It also gives them the impression that the university cares about them as individuals and wants them to succeed, despite their initial shortcomings.

Another feature that this course has to offer is its small class size. Each section of the course is capped at about 30 students, so unlike Calculus 1, it is easy for students to get personal attention from the instructor. As mentioned in the previous section of this chapter, this is a desirable trait for first generation students, or any student who is having difficulty adjusting to the large lectures that are typical at a large research institution. Also, most of the sections have had more in common with recitations than lecturers, as students had ample time to work on practice problems and homework during class. Having an additional semester to get adjusted to college life while still being able to enjoy the benefits of a high school-style course could make the difference between a student staying in school or dropping out.

Since the Precalculus students are spending a good deal of time in class working through the algebra and trigonometry material they may have struggled with before, the foundation is laid for their future success in mathematics. However, this course is more than an algebra and trig review. The problems are different from traditional algebra/trig problems in that they have students manipulate expressions into convenient forms, rather than solve for a variable (though these sorts of problems are given as well). This skill is a common weak point for Calculus students, and will prepare them for the sorts of problems they will encounter later in their mathematical career.

Finally, the calculus review at the end of the course can serve as a confidence boost for those students who haven't seen the material before. As mentioned in the previous section, feeling like you are the only one who hasn't seen the material before can be intimidating.

Upon completing Precalculus, a student should have the mathematical foundation as well as the confidence to do well in the beginning of Calculus 1, should they choose to take it again.

Research Questions

In summary, the researcher poses the following questions to guide their investigation:

1. In large lectures, it is difficult or impossible to teach to students as individuals. Can data mining break the Calculus 1 large lecture course into different groups to allow the lecturer to address each type of student? If so, what would be the nature of these groups?

2. For Calculus 1, is it possible to quickly identify students who are in need of extra assistance? Is Bailout Precalculus a good way of helping these students?

3. If Bailout Precalculus is effective at changing the trajectory of students' mathematical and educational progress, why is it effective?

Limitations and Challenges

There are several limitations that constrain the generalization and accuracy of the results of this study. The most serious limitation is that all of the subjects of this study were students of Kansas State University. This school is a large, research institution in the Midwestern United States, so the number, type, and "personality" of the clusters may be different if calculated using data from another type of school. Even within Kansas State, students of Calculus 1 may not be representative of the student body as a whole, for the majority of them are freshman engineering majors.

Since the aim of this study is to find a way to reach out to struggling students, the researcher feels it is necessary to perform the cluster analysis early in the semester (immediately after the first exam). Due to this, the number of assignments that can be used in the analysis is limited, usually to a few homeworks and online homeworks, attendance, and a midterm. Also, unlike in Manspeaker's work, a mechanism is not in place for the researcher to obtain the scores for individual exam questions, so the variance among the students' scores will be constrained.

Since the clustering algorithms only take into account quantitative data, interviews with students are necessary to discover the qualitative characteristics of the groups. However, these interviews can lead to some problems. To incentivize students to participate in the interviews, $15 compensation will be offered for the 10-30 minutes of their time that it takes to answer the questions. This has the potential to disproportionately attract certain types of students (such as lower income students) over others, thus skewing the researcher's assessment of the personality of a cluster. Furthermore, since the students need only complete the interview to receive the reward, some students may give short or incomplete answers to finish as quickly as possible. The students answers may also be influenced by the fact that they would be speaking with the researcher, who is known to be a member of the mathematics department. Finally, if a cluster is small enough, it may be difficult to recruit its members for interviews due to scarcity. This could result in an inaccurate representation of that cluster's characteristics.

The above issues, except for the last one, also apply to the interviews conducted with Precalculus students. Small population size is an issue for Precalculus, since each section has just 30 students, and the course has only been offered for four semesters (each Fall since 2013). Selection bias is also a factor, as it may be difficult to reach students who have left the university for interviews.

Summary

The primary goals of this study are to understand what sorts of students are enrolled in large lecture Calculus 1 courses, if it is possible to differentiate instruction in such a class, and how intervention affects struggling and unprepared students. Given the size of Calculus 1 at Kansas State University, data mining methods must be used to analyze similarities between the students. This, combined with interviewing select students, will hopefully yield a sensible partition of the class that will be amenable to differentiated instruction.

The researcher believes that Bailout Precalculus rescues students with weak mathematical backgrounds, or those who are simply unaccustomed to typical large lecture courses, from starting a trend of failure by giving them a second chance to pass their first college math course. Through qualitative analysis of students' reactions and experiences, the researcher hopes to understand the impact of this unique class and how it has shaped these students' futures.

Chapter 2

Literature Survey

This chapter is a survey of the relevant background information on differentiated instruction, data mining techniques, and "bailout" courses used in this study. While there is a wealth of information on differentiated instruction and clustering algorithms in the literature, there has been relatively little work published on courses such as Bailout Precalculus. Thus, the majority of this chapter will be spent on topics related to the first two research questions posed in Chapter 1. Since this is a literature survey, all material introduced in this chapter is well known and has previously been published by other authors.

Educational Data Mining

What exactly is data mining? According to the textbook "Introduction to Data Mining" by Tan, Steinbach, and Kumar, "Data mining is the process of automatically discovering useful information in large data repositories" Tan et al. (2005). Prior to the era of electronic computers, it would be difficult or impossible for an individual to collect, sift through, and create models from vast amounts of data. With the invention and proliferation of the computer, it became possible to find and use this information to discover previously unknown patterns. Thus, around the 1960's Lovell (1983), data mining (or data fishing, as it was referred to in that time) began to emerge as a field of study.

Initially, data mining was used primarily for scientific and business applications. However, large data sets also appear in the world of education, especially with the increase in university enrollment and average class size. In the early 2000's, data mining was first applied in education by analyzing students' usage of computer based learning tools. In 2008, the first annual International Conference on Educational Data Mining was held, and in 2009, the Journal of Educational Data Mining was created EDM (2008). Today, a wide variety of educational questions are analyzed through the lens of educational data mining, such as prediction of student performance, grouping students, concept maps, or providing feed back for supporting instructors Romero and Ventura (2010).

Data Mining Techniques

The techniques used in Educational Data Mining are not specific to educational applications - they can be used on any large set of data. The following several sections are a description of the techniques the researcher used in their study. While these techniques are not commonly used, they were appropriate given the nature of the data the researcher analyzed.

Formatting and Preprocessing the Data

For each semester of Calculus 1, there is a total of several hundred students enrolled in the course. Each student can be viewed as a vector x contained in $\mathbb{R}^N$, where N is the number of assignments, exams, and attendance values each student has. Note that while all assignment values are rational, there is nothing to be gained by considering it as a vector in $\mathbb{Q}^N$. Since one of the objectives of this study is to understand which students are similar to each other and which are different, a notion of distance must be defined.

There are uncountably many different ways one can define a metric, i.e. a distance on $\mathbb{R}^N$, but for the purposes of this study, the standard metric will suffice. This is the typical

Euclidean distance between two points, and is defined for all $x, y \in \mathbb{R}^N$ by

$$d(x,y) = \sqrt{(x_1 - y_1)^2 + ... + (x_N - y_N)^2},$$

where $x = (x_1, x_2, ..., x_N)$ and $y = (y_1, y_2, ..., y_N)$.

In addition to considering the distance between two points, it is also important to know how to measure the correlation of two attributes. The correlation is thought of as "how linear" the relationship between two attributes is. This value is always between -1 and 1, where values close to 1 suggest a strongly positive linear relationship, and values near -1 are indicative of a strongly negative linear relationship. For attributes with n components, say u and v, the correlation is defined as

$$corr(u,v) = \frac{cov(u,v)}{\sigma_u * \sigma_v},$$

where

$$cov(u,v) = \frac{1}{n-1}\sum_{j=1}^{n}(u_j - \bar{u})(v_j - \bar{v}),$$

is said to be the covariance of u and v, $\bar{u}$ and $\bar{v}$ are the means of u and v respectively, and σ_u and σ_v are the standard deviations of u and v, which is defined as

$$\sigma_u = \sqrt{\frac{1}{n-1}\sum_{j=1}^{n}(u_j - \bar{u})^2}.$$

It is clear why the distance between two points is important, but why must one consider correlation? Later in this chapter, it will be demonstrated that there are difficulties in dealing with high dimensional data sets. In order to reduce the dimension of the data without losing a significant amount of information, one must discard data that is strongly correlated with data that is being kept. If the two are strongly correlated, one can draw similar conclusions from them, so keeping both sets is redundant.

Next, it is necessary to normalize the data. Certain assignments in Calculus 1 have different point values than others. For example, a student can earn a maximum of 10 points on a homework assignment, but can earn up to 100 points on an exam. To prevent assignments with higher maximum scores from having undue influence on the clustering, each assignment will be normalized. Let x be a student's score on a particular assignment. Their normalized score x' is defined in the following way:

$$x' = \frac{x - \bar{x}}{\sigma_x},$$

where $\bar{x}$ is the mean and σ_x is the standard deviation of the assignment among all students. Transforming the scores in this way results in a mean of 0 and a standard deviation of 1 for all assignments.

Clustering

After normalizing the students' scores, it is time to break them into different clusters. One of the most commonly used methods of partitioning a data set into clusters is called The K-Means Algorithm, or just K-Means. The algorithm is as follows Tan et al. (2005):

1. Select K points as initial centroids (centers of the clusters).
2. **repeat**
3. Form K clusters by assigning each point to its closest centroid.
4. Recompute the centroid of each cluster.
5. **until** Centroids do not change.

Note that the number of clusters needs to be decided before the algorithm begins. This can be chosen by using expert knowledge (white-box methods), or by using another clustering algorithm called Agglomerative Nesting (or AGNES) which will be described later in this chapter. Also, the initial K centroids can be chosen at random, or can be selected to be far apart from one another to obtain more distinct clusters.

We use the Euclidean distance to determine which centroid is closest to a given data point. To determine which centroids are optimal, we consider a quantity called the Sum of Squared Error (SSE). Computing the optimal clustering is equivalent to minimizing the SSE (the objective function of this minimization problem). The SSE is given by the following formula:

$$SSE = \sum_{j=1}^{K} \sum_{x \in C_j} d(x, c_j),$$

where c_j is the centroid of the jth cluster.

One can also define the SSE of a particular cluster by only summing over the points x in that cluster. By minimizing this value, the optimal centroid of that cluster can be found. As it turns out, if the distance used is Euclidean, and the objective function is the SSE, the optimal centroid is simply the arithmetic mean of the points in the cluster.

Proposition The SSE of a cluster is minimal if its centroid is the arithmetic mean of the points in the cluster.

Proof Let $c = (c_1, ...c_N)$ be the centroid of the cluster C. We will show that if the SSE is minimized, then c will equal the mean. Considering the SSE as a function of c, we can find the gradient, and set it equal to 0 to find the local extrema. Since the SSE is symmetric with respect to its coordinates, it suffices to examine the partial derivative in one coordinate, say the jth. This calculation proceeds as follows:

$$\frac{\partial SSE}{\partial c_j} = \frac{\partial}{\partial c_j} \sum_{x \in C} \sum_{k=1}^{N} (c_k - x_k)^2 = 0$$

$$\sum_{x \in C} \sum_{k=1}^{N} \frac{\partial}{\partial c_j} (c_k - x_k)^2 = 0$$

$$\sum_{x \in C} 2(c_j - x_j) = 0$$

$$\sum_{x \in C} 2c_j - \sum_{x \in C} 2x_j = 0$$

$$\sum_{x \in C} c_j = \sum_{x \in C} x_j$$

$$Nc_j = \sum_{x \in C} x_j$$

$$c_j = \frac{1}{N} \sum_{x \in C} x_j = \bar{x_j}.$$

Since $(\bar{x_1}, ..., \bar{x_N}) = \bar{x}$, $c = \bar{x}$, as desired. ■

In general, it is easier to prove theorems about means than medians, making K-Means a powerful tool theoretically, as well as practically. However, one drawback to using K-Means is that it is susceptible to being unduly influenced by outliers. In educational data sets, outliers are a real possibility (for example, a student who turns in none of the assignments and does very poorly on the exams). To account for this, a modified version of K-Means called K-Medoids (also called Partitioning About Medoids, or PAM) will be used, which uses points within the data set itself as the centers for the clusters.

In PAM, rather than using the geometric locations of the data points, their dissimilarity (or the Euclidean distance between points) is used. The PAM algorithm seeks to minimize the average dissimilarity between all possible pairs of points, which replaces the SSE as the objective function. This changes makes PAM different from K-Means, and thus changes the implementation algorithm.

PAM takes place in two stages: BUILD and SWAP. In the BUILD phase of PAM, the initial medoids are strategically chosen to make the SWAP phase converge quickly on a solution. As an algorithm, BUILD looks like the following:

1. Find the point in the data set for which the sum of dissimilarities with all other points is minimized. This point will be the first medoid (call it m).

2. Let x and y be two points not equal to m. Let D_x be the dissimilarity between x and the current medoid, and let $d(x, y)$ be the dissimilarity between x and y. If $D_x > d(x, y)$, then y would make a good medoid "from x's perspective", so x will contribute to y being the next medoid. This is expressed as a term C_{xy} =max$(D_x - d(x, y), 0)$.

3. Repeat step 2 for each other point x, and then add all of the terms together to compute the total "votes" for y as the next medoid: $G_y = \sum_x C_{xy}$.

4. Choose as the next medoid the point that has the largest G_x.

5. Repeat steps 2-4 until K medoids have been selected.

After the BUILD phase, K different medoids have been chosen, but they might not be the optimal choices. This is what the SWAP phase is for - medoids are swapped with other points, and the total dissimilarities are compared to see if the new medoid is superior to the old. Specifically the SWAP algorithm looks like this:

1. Let m be a current medoid, and x be a non-medoid point in the data set that m potentially may be swapped with. Let y be another non-medoid point that is not equal to x. Let C_{mxy} be y's contribution or "vote" to swap x and m. Negative values of C_{mxy} will push the decision towards swapping, whereas positive values will contribute towards keeping m as a medoid.

   (a) If y is more dissimilar from both m and x than another medoid, C_{mxy} is 0.

   (b) If m is the closest medoid to y, then there are two possibilities:

      i. The first case is where y is closer to x than the second closest medoid, which implies that $d(x, y) < E_y$ (E_y is the distance from y to the second closest medoid). In this scenario $C_{mxy} = d(x, y) - D_y$.

      ii. The other possibility is if y is further away from x as the second closest medoid, or in other words $d(x, y) \geq E_y$. In this case $C_{mxy} = E_y - D_y$.

   (c) If y is further away from m than other medoids, but closer to x than any medoid, then $C_{mxy} = d(x, y) - D_y$, where it should be noted that $D_y \neq d(y, m)$, since m is not the closest medoid to y.

2. Repeat the first step for all y not equal to m or x, then sum the results: $S_{mx} = \sum_y C_{mxy}$

3. Find the pair (m, x) where T_{mx} is negative and T_{mx} is minimal among all possible pairs. Swap m and x.

4. Repeat steps 1-3 until all the T_{mx} are 0 or positive. At that point, the SWAP algorithm terminates.

Kaufman and Rousseeuw (1990)

Note that while PAM is guaranteed to converge to a locally optimal solution, it is possible that this solution is not globally optimal, like it is in K-Means. However, the chance of this occuring is minimal because of the BUILD algorithm (it selects good starting medoids). Another drawback to PAM is that the preceding algorithms were more complicated than those of K-Means. These downsides are an acceptable price to pay in our situation, as by using PAM for clustering we have minimized the impact of outliers, and have as a representative element of each cluster a member of our data set.

The observant reader may note that we needed to know the number of clusters, K, in advance for both K-Means and PAM. How do we know how many clusters there should be for a given data set? The researcher could simply make an educated guess based on the nature of the data, or try several different K values and see which one seems to make the most sense, but doing so would be arbitrary to a certain degree. Instead, another clustering algorithm, Agglomerative Nesting (AGNES), will be used to choose K.

What is AGNES, and how will it help us determine the appropriate number of clusters? First, let us look at an algorithm for AGNES:

1. For all pairs of distinct points in the data set (x, y), compute the Euclidean distance between them $d(x, y)$.

2. Find the pair that minimizes this distance. These points are now considered to be in the same cluster.

3. Repeat steps 1-2 for the second closest pair, then the third closest and so on, until all points are in the same cluster.

Initially, this algorithm seems strange, as the end result will have all of the points in one cluster. What is done in practice is that the AGNES algorithm is prematurely terminated when the distance between the next pair of points reaches a certain threshold. Then, the number of clusters is counted, and this is the value of K that is used either in K-Means or PAM.

Since it may be difficult to determine an appropriate cutoff point in advance, this process is often done using a diagram called a dendrogram. An example is given below: How does

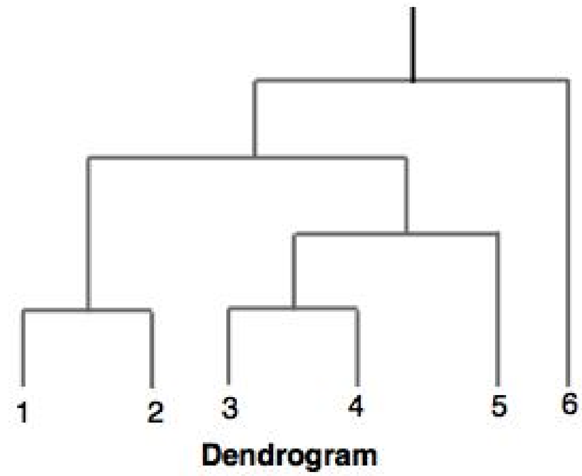

Figure 2.1: *Example of a dendrogram*

one interpret this diagram? In the above example, our data set has 6 points. It appears that the distance between 1 and 2 is the roughly the same as the distance between points 3 and 4, so these are pairs form the first two clusters. Then we observe that point 5 and the 34 cluster are the next closest, so those are grouped together. After this, cluster 12 and cluster 345 are grouped together, so at this point we have a cluster with points 1-5, and 6 in its own cluster. For the last step, 6 joins the other points and the entire data set is a single cluster. If the abridged version of AGNES were being carried out, it might be reasonable for the user to stop the process with clusters 12,345,6 or perhaps 12345,6, meaning that the "natural" number of clusters K for PAM or K-means would be 2 or 3.

One may wonder why we need to bother with the significantly more complicated PAM if AGNES is itself a clustering algorithm. The reason for this lies in the fact that AGNES is an example of a "nearest-neighbor" algorithm. This means that the cluster a point x belongs to depends on its neighbors, or the closest points to x. The problem with this is that it makes the composition of clusters unstable. For example, suppose x and y have a distance of 1, but y and z have a distance of 1.1, and that (x, y) is the closest pair in the data set. Then a small of shift of about 0.1 to y could cause it to pair up with z first, and since it was the first grouping in the algorithm, it could potentially affect the composition of all of the clusters.

Such a shift is realistic with educational data sets, where someone could accidentally hit the incorrect number on the keyboard when entering grades, for example. Fortunately, even if the composition of the clusters is unstable, it can be shown that the number of clusters before a given cutoff point is stable under small perturbations of the points.

Dimension Reduction

One obstacle to obtaining stable clusters is that of high dimensional data. Why is this an issue? Consider for example the interval [-1,1] a 1 - dimensional ball of radius 1 centered at the origin on the real line. The part of this interval from -0.9 to -1 and 0.9 to 1, i.e. the outer 10%, takes up 10% of the length of the figure.

Figure 2.2: *The outer 10% of the interval [-1,1]*

Now let us move up a dimension. In two dimensions, the outer 10% of the ball has area $A_{10\%} = \pi 1^2 - \pi 0.9^2 = 0.19\pi$, meaning it has 19% of the total area of the circle - more than in the 1 dimensional case.

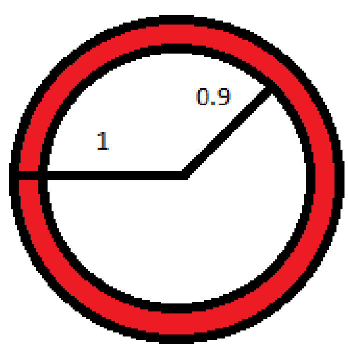

Figure 2.3: *The outer 10% of a disk of radius 1.*

This trend continues as the dimension increases. In general, for an n-dimensional ball of radius 1, the fraction of the area in the outer 10% is

$$P_{10\%} = \frac{\pi 1^n - \pi 0.9^n}{\pi 1^n} = 1 - 0.9^n,$$

which tends to 1 as n grows large. This result, which initially seems quite strange, is only counterintuitive because we are used to working in a low number of dimensions.

Why is this important? Suppose that the center of an n- dimensional ball of radius 1 is the origin, and that this point represents the center of a cluster of data. Now, if a small change is made in the coordinates of this point, but the radius remained 1, the entire disk would shift as well. This can be a serious problem, as points that were originally in the periphery of the original region would not be in the transformed region. Also, as demonstrated in the calculations above, a significant percentage of the points lie in the outer rim of the disk. Thus a tiny change in the cluster's center can cause a large number of points to switch clusters, making the clustering unstable.

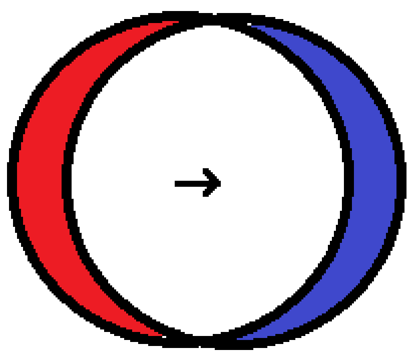

Figure 2.4: *The points in red are lost from the cluster, and the points in blue switch into the cluster.*

In this work, the data sets are up to 17 dimensional, resulting in $1-0.9^{17} \approx 0.83$ of the area being in the outer 10% of a ball. How can a stable clustering be formed with this data? One way is for the researcher to select the dimensions they deem most important in the data and throw out the other dimensions. In the case of educational data, this amounts to selecting the assignments that seem to provide the most variation and discounting the rest. This type of dimension reduction is an example of "White Box Analysis", where the researcher uses expert knowledge to select the most impactful parts of the data. However, this approach removes the potential for finding unexpected results and relationships in the data, as well as introducing the potential for human error. As such, a “Black Box”, algorithm-based approach is more appropriate for this study.

The researcher chose to use Singular Value Decomposition (SVD) to reduce the dimension of the data. In SVD, the goal is to find a projection $\mathbb{R}^D \to \mathbb{R}^d$, where D is the original, large dimension and d is some smaller dimension (typically 2-6 for many applications). We also want this projection to preserve as much variance among the data as possible to find interesting results from the data. For example, if every student attended class on the same day, including this information would not help in clustering, whereas something highly varied like exam scores would.

Let us find such a projection. The first step in doing so is to verify the following result:

Proposition 1: Let $\mathbf{A}$ be a $n \times D$ matrix where the column vectors are standardized. Suppose $x \in \mathbb{R}^D$ has magnitude 1. Then the projection $\mathbb{R}^D \to \mathbb{R}$ defined by $a \mapsto a \cdot x$, where a is a row of $\mathbf{A}$ that maximizes the variance of $a \cdot x$ occurs when x is an eigenvector of of $\mathbf{A}^T\mathbf{A}$ corresponding to the largest eigenvector λ_1.

Proof: Since each column of $\mathbf{A}$ is standardized, $E(a \cdot x) = 0$. This implies that the variance is

$$V(a \cdot x) = E((a \cdot x)^2) - E(a \cdot x)^2 = E((a \cdot x)^2) = |a \cdot x|^2.$$

The last term is equivalent to

$$(\mathbf{A}x) \cdot (\mathbf{A}x) = (\mathbf{A}x)^T(\mathbf{A}x) = x^T\mathbf{A}^T\mathbf{A}x,$$

so maximizing the variance over all x where $|x| = 1$ is equivalent to maximizing $x^T\mathbf{A}^T\mathbf{A}x$.

This can be accomplished with Lagrange multipliers, since matrix multiplication is smooth and our domain, S^D, is compact. Since we have one constraint, we only need one multiplier;

call it λ. We want to find the critical points of

$$x^T \mathbf{A}^T \mathbf{A} x - \lambda(|x| - 1).$$

To do this, name the entries of $\mathbf{A}^T\mathbf{A}$ as (α_{ij}). Using this and the definition of matrix multiplication, we can write the above expression as

$$\sum_{ij} \alpha_{ij} x_i x_j - \lambda \sum_{i=1}^{D} x_i^2 - \lambda.$$

Note that since $|x| = 1$, $\sqrt{\sum_{i=1}^{D} x_i^2} = \sum_{i=1}^{D} x_i^2$. Since $\mathbf{A}^T\mathbf{A}$ is symmetric for any matrix $\mathbf{A}$, we can rewrite our Lagrangian as

$$2 \sum_{i<j} \alpha_{ij} x_i x_j + \sum_{i=1}^{D} \alpha_{ii} x_i^2 - \lambda \sum_{i=1}^{D} x_i^2 - \lambda.$$

Now the expression is ripe for partial differentiation with respect to each variable x_i:

$$\frac{\partial}{\partial x_i}(2 \sum_{i<j} \alpha_{ij} x_i x_j + \sum_{i=1}^{D} \alpha_{ii} x_i^2 - \lambda \sum_{i=1}^{D} x_i^2 - \lambda) = 2 \sum_{j=1}^{D} \alpha_{ij} x_j - 2\lambda x_i,$$

and for λ:

$$\frac{\partial}{\partial \lambda}(2 \sum_{i<j} \alpha_{ij} x_i x_j + \sum_{i=1}^{D} \alpha_{ii} x_i^2 - \lambda \sum_{i=1}^{D} x_i^2 - \lambda) = - \sum_{i=1}^{D} x_i^2 + 1 = -|x| + 1.$$

Setting these derivatives equal to 0 yield the equations $|x| = 1$ and

$$\sum_{j=1}^{D} \alpha_{ij} x_i = \lambda x_i.$$

Notice that this last equation written in matrix form is $\mathbf{A}^T\mathbf{A}x = \lambda x$, meaning that λ is an eigenvalue of $\mathbf{A}^T\mathbf{A}$. Observe that

$$V(a \cdot x) = |\mathbf{A}x| = \sqrt{(\mathbf{A} \cdot x)(\mathbf{A} \cdot x)} = \sqrt{x^T\mathbf{A}^T\mathbf{A}x} = \sqrt{x^T\lambda x} = \sqrt{\lambda x^T x} = \sqrt{\lambda},$$

so we should select the eigenvector x that has the largest eigenvalue to maximize the variance. ■

If we were seeking a 1-dimensional projection, then the projection constructed above would be optimal. However, reducing high dimensional data down to just 1 dimension would be excessive, so we need to construct a projection that lowers the dimension of our data to any desired dimension. We shall do so by continuing the construction above.

Proposition 2: Let v_1 be the eigenvector of $\mathbf{A}^T\mathbf{A}$ with the largest eigenvalue λ_1 found in the previous proposition. Then the vector x that maximizes $V(a \cdot x)$ subject to the constraints $|x| = 1$ and $v_1 \cdot x = 0$ (i.e. x is perpendicular to v_1) is the eigenvector of $\mathbf{A}^T\mathbf{A}$ with the second largest eigenvalue λ_2.

Proof: Since the objective function in this proof is the same as the last, we simply need to find the critical points of the Lagrangian

$$x^T\mathbf{A}^T\mathbf{A}x - \lambda(|x| - 1) - \mu(v_1 \cdot x),$$

with μ being the second Lagrange multiplier for our new constraint $v_1 \cdot x = 0$. Writing this using sums, we obtain

$$2\sum_{i<j} \alpha_{ij}x_i x_j + \sum_{i=1}^{D} \alpha_{ii}x_i^2 - \lambda \sum_{i=1}^{D} x_i^2 - \lambda - \mu \sum_{i=1}^{D} v_{1i}x_i,$$

using the same reasoning as the preceding proposition. Taking the same derivatives as before, along with the μ derivative, and setting them equal to 0 we have the equations $|x| = 1$, $v_1^T x = 0$, and

$$2\sum_{j=1}^{D} \alpha_{ij}x_j - 2\lambda x_i - \mu v_{1i} = 0 \implies 2\mathbf{A}^T\mathbf{A}x - 2\lambda x - \mu v_1 = 0.$$

Using this information and multiplying by v_1^T on the left produces

$$0 = 2v_1^T\mathbf{A}^T\mathbf{A}x - 2\lambda v_1^T x - \mu v_1^t v_1 = 2v_1^T\mathbf{A}^T\mathbf{A}x - \mu = 2(\mathbf{A}^T\mathbf{A}v_1)^T x - \mu = 2(\lambda v_1)^T x - \mu = -\mu.$$

This implies that $\mu = 0$, so we get

$$2\mathbf{A}^T\mathbf{A}x - 2\lambda x = 0$$

implying that once again x is an eigenvector of $\mathbf{A}^T\mathbf{A}$ with eigenvalue λ. As before, maximizing the variance is equivalent to maximizing $\sqrt{\lambda}$, but since we can't use λ_1, we must use the eigenvector v_2 with the second largest eigenvalue λ_2. ■

Note that in the above propositions, we implicitly assumed that all of our eigenvalues were real, as to exploit the ordering on the reals. There was not problem in doing this, as symmetric matrices like $\mathbf{A}^T\mathbf{A}$ always have real eigenvalues. Also observe that there was nothing special about doing this for 2 dimensions - we can repeat this process until we have an orthonormal basis $(v_1, v_2, ..., v_D)$ for our original space of data points. Now, if our original matrix A is of full rank (which is very likely for educational data sets), then we can define the following basis vectors

$$u_i = \frac{\mathbf{A}v_i}{\sqrt{\lambda_i}},$$

where v_i and λ_i are defined as above. Note that

$$|u_i| = \frac{|\mathbf{A}v_i|}{\sqrt{\lambda_i}} = \frac{\lambda_i}{\lambda_i} = 1$$

for each i, and for $i \neq j$, $u_i^T u_j = 0$, so the u_i form an orthonormal subset of $\mathbb{R}^n$ for $n \leq D$. Though this set does not form a basis of $\mathbb{R}^n$ in general, it is a piece of the SVD which can form such a basis.

The definition of a (Full) Singular Value Decomposition is the following: Let $\mathbf{A}$ be a $n \times D$ matrix of full rank. Let $\mathbf{U}$ be the matrix with u_i as the ith row and $\mathbf{V}$ be the matrix with v_i as the ith row. Also let $\mathbf{S}$ be the diagonal matrix with $\sqrt{\lambda_i}$ (also known as the singular values) as the ith entry. Then the Full Singular Value Decomposition of $\mathbf{A}$ is given by

$$\mathbf{A} = \mathbf{U}\mathbf{S}\mathbf{V}^T,$$

which means we have the following commutative diagram:

$$\begin{tikzcd}
\mathbb{R}^D \arrow[r, "\mathbf{A}"] \arrow[d, "\mathbf{V}^T"'] & \mathbb{R}^n \\
\mathbb{R}^D \arrow[r, "\mathbf{S}"] & \mathbb{R}^D \arrow[u, "\mathbf{U}"]
\end{tikzcd}$$

How does one use the SVD for dimension reduction? First, note that the columns of the matrix $\mathbf{V}$ are the new basis vectors for our data set. Unlike the standard basis for Euclidean space, the entries of these vectors place a weight on the value of each dimension/assignment that will maximize the variance. The diagonal matrix $\mathbf{S}$ contains the singular values, indicating how much variance is preserved by each basis vector v. By construction, the singular values decrease down the diagonal. Finally, the columns of $\mathbf{U}$ represent the new coordinates of the data points with respect to the new basis. However, since $\mathbf{V}$ is $D \times D$, we have not yet reduced the dimension of our data. This leads us to the Truncated Singular Value decomposition.

The Truncated Singular Value decomposition of a full-rank matrix $\mathbf{A}$ is given by the following equation:

$$\mathbf{A}_d = \mathbf{U}_d\mathbf{S}_d\mathbf{V}_d^T,$$

where $\mathbf{U}_d$ has the rightmost $D - d$ columns removed, $\mathbf{S}_d$ only keeps the first d of its diagonal matrix (and is now $d \times d$) and $\mathbf{V}_d$ only has as columns $v_1, ..., v_d$. Since the size of this new matrix is only $n \times d$, $\mathbf{A}_d$ is clearly a different matrix from $\mathbf{A}$. By construction, $\mathbf{A}_d$ is the best rank d approximation of $\mathbf{A}$, meaning it preserves as much variance as possible using a d-dimensional basis.

One final issue remains - since a truncated SVD can be constructed for any d value between 1 and D, how do we know which one to select? The answer to this question can be found by examining a graph of the singular values of A. Take for example the singular values plotted below:

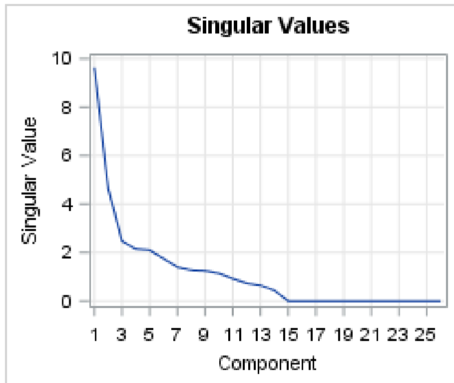

Figure 2.5: *A plot of the value of the singular values vs. their position in* $\mathbf{S}$.

In this figure, notice that while the values decrease sharply initially, they level off after the third value. This is what is called the "elbow" of the singular values, and its location indicates an appropriate dimension for truncation. After the elbow, observe that the values begin to level off, which means that there are diminishing returns to the variance for including more dimensions of the data. Thus, for this example, it's clear that $d = 3$ is appropriate.

There are two issues with choosing d in this manner. The first is that for some data sets, particularly the messy data sets encountered in education research, it may be difficult to determine where the elbow of the singular values is. Consider for example the following singular plot: Unlike the previous set of singular values, this one does not have a clear elbow,

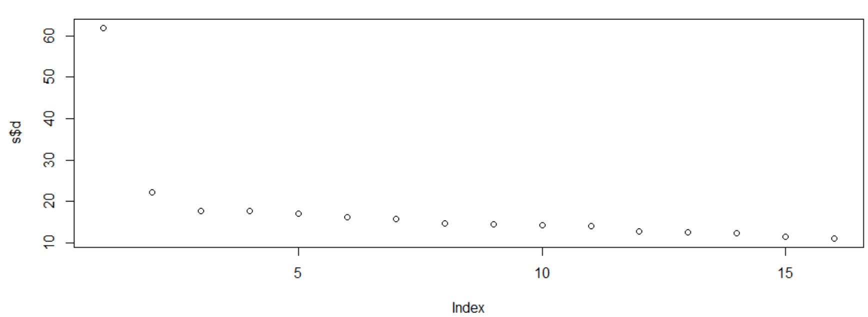

Figure 2.6: *This time, it's not as clear where the "elbow" should be.*

making the precise d to truncate to difficult to determine. In practice, experimenting with several different d values and comparing the results of clustering the truncated data can lead to a suitable choice. The other problem with this method is that, like AGNES, it is ultimately subjective since finding the elbow relies on a human decision. However, since the choice is either obvious or can be determined experimentally as described above, this problem's impact is limited.

Retention and Student Difficulties

Retention is a serious issue in modern STEM education. Between 1987 and 1991, 44.1% of students who began as STEM majors switched to non-STEM majors by the time they were seniors. In mathematics specifically, the situation is worse with 62.6% of students switching out of any sort of STEM major Seymour and Hewitt (1997). These percentages are even higher for female students, at 52.4% and 72.3% respectively.

What is the cause for this attrition? According to the book "Talking About Leaving" by Elaine Seymour and Nancy Hewitt, there are many reasons why students give up on their original path. The first cause is derived from the students' original motivation for choosing a STEM major. Among the students who choose to stay in their field, common motivations include intrinsic interest in the subject, materialistic concerns, and the influence of others (in descending order of frequency among those surveyed). These are also important reasons for those who switched, but the order of importance is different: influence of others, materialism, and being skilled at math/science in high school were the reasons why they began in their major. Furthermore, the influence of others was mentioned significantly more often among female students who left, whereas being skilled in high school was stated more often by male students Seymour and Hewitt (1997). This data suggests that external influence (quite often parental) and the lure of a large salary push students who may not be as enthusiastic about a technical field into STEM.

It is often the case that expectations and standards in high school are far less demanding than those placed upon students in college. About 40% of those surveyed in "Talking

About Leaving" cited inadequate high school preparation as one of the primary reasons they switched. Also, as mentioned above, many students who ended up switching saw themselves as strong students in the sciences prior to enrolling at university. One student stated that "I was seen as a very, very good student in high school because I was a natural at math. But when I had to retake both semesters of calculus, I realized that my math preparation hadn't been all that good." These deficiencies become more prominent when the students compare themselves to the other students in their classes: "There was just so much that I didn't know... I felt like I wasn't even in the same game as everyone else." Seymour and Hewitt (1997) Furthermore, even when students are recognized as being unprepared for university level technical courses (especially mathematics), they are often "sentenced" to taking remedial courses for a semester or a year, which can be demoralizing.

These feelings of not measuring up in an actual or perceived competitive environment also contributes to student attrition. Students become resentful of those who have already seen the material previously: "We know quite a number of people who have taken A.P. physics who are in this class and they shouldn't be. This is a first class... and the professor went so quickly for them, the rest of us got lost". Seymour and Hewitt (1997) This attitude becomes even worse in courses (like Calculus 1) that are considered to be "weed-out" courses. Since the students believe that teacher is trying to get many of them to drop the course or leave a certain major, they often become afraid or resentful of the professors themselves, thus reducing the chance of these students seeking them out for assistance.

The difference between high school and college classes also manifests in both the pace of the curriculum and the amount of work assigned. In college, students have to master the same amount of material in one semester that would have been covered in an entire school year during high school. About one third (34.9%) of students who switched out of STEM cited this increased pace, as well as a demanding workload, as a primary reason they left the sciences Seymour and Hewitt (1997). One student stated that

"I guess what makes it hard is both the intellectual grasp needed, and the tremendous volume of work- especially coming from high school. It's just like drinking from a fire hose - all this stuff you've never seen before in your life. And there's none of the personal attention you're used to in high school. The pace is just incredible. You either sink or swim: and a lot of them end up sinking." Seymour and Hewitt (1997)

As this student mentioned, lack of personal attention is also a potential barrier to learning, which is a very common feature in large lecture courses such as Calculus 1. Though the students have access to their recitation instructors, these sections can themselves be too large for personal attention (about 40 students each in the case of K-State's Calculus 1 course). This, combined with having to master more difficult material twice as quickly as in high school is a recipe for disaster for some students.

This increased pace forces the lecturer to go through the material quickly, leaving little time for examples. One student mentions "It's the lack of examples, especially complex examples... One important thing the professor should do in class is to supplement the book with lots of examples. But that doesn't happen." Seymour and Hewitt (1997) Going through particular applications of the material is often relegated to recitation, but depending on available time, it may happen that a particular type of problem or concept is never covered.

There are many issues with large lecture courses in college, but problems with calculus often begin in high school. In the MAA publication "The Role of Calculus in the Transition From High School to College Mathematics", the AP Calculus exam and its effects are discussed. Originally, taking calculus in high school was very uncommon, with most students taking "Analytic Geometry" their first semester in college, which is closely related to modern high school precalculus. The advanced placement test was introduced as a way to allow a small number of advanced students to get ahead in mathematics, but over time it had the effect of accelerating the high school mathematics curriculum so more students could enroll in calculus their senior year. In recent years, many view taking AP Calculus as a way to be competitive in college admissions, so many students enroll in the class despite not being prepared for it. Rosenstein and Ahluwalia (2016)

Furthermore, many of the calculus classes that high school students take aren't AP courses. A study at Rutgers found that 43.3% of students who had taken a full year calculus course in high school took a non-AP course. These courses vary wildly in content and often give students a false sense of security as they enter college.

The students who don't do well enough on the AP exam to gain credit for Calculus 1 (usually a score of 1 or 2 on the AP AB exam) were more likely to have a negative rationale for taking AP Calculus than those who passed the exam. For example, 30.3% of these students claimed to have taken the AP course so they wouldn't have to take math during college, as opposed to 5.1% of those who got a 4 or 5 on the AB exam (the scores required for credit at Rutgers). Also, only 52% of these students said they enjoy taking challenging math courses while 79% of the ones who passed the exam said they did. Overall, these students are not only more likely to be enrolled in Calculus 1 than their peers who passed the exams - they are also more likely to be unenthusiastic about it.

Calculus Intervention

It is uncommon for a university to have a drastic intervention strategy for Calculus I like the one described in this work. The University of Michigan is one of the exceptions - as early as 2004 they had a calculus intervention program very similar to Kansas State's Bailout Precalculus course. Koch and Herrin (2006) Earlier, the researchers at Michigan observed a relationship between a student's grade on their first attempt at Calculus 1 and their likelihood of graduating university within six years:

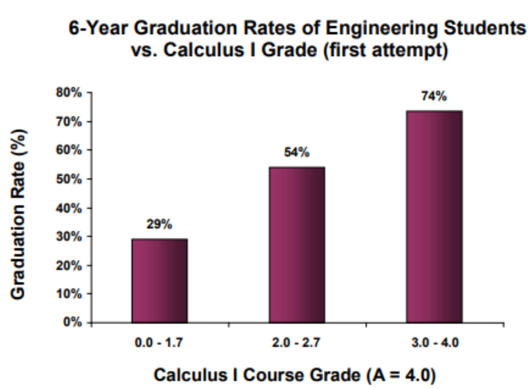

Figure 2.7: *The higher the grade in Calculus 1, the more likely a student is to graduate within 6 years. Koch and Herrin (2006)*

Thus, Koch and Herrin launched a study of their own Bailout Precalculus course, MATH 110. The format of this course is extremely similar to that of Kansas State's Precalculus course, in that students who do poorly on the first exam are given the opportunity to switch into Precalculus and have their grades in Calculus 1 expunged. The only differences were that the students in danger of failing Calculus 1 attended an informational session to inform them about the bailout option, and that the course was run as an independent study do to scheduling issues. Koch and Herrin (2006)

Of the 397 engineering students enrolled in Calculus 1 in the Fall of 2004, 57 received a C- or lower on the first exam - the cutoff for being invited to participate in the bailout. Of these 57, 25 chose to enroll in the bailout (43.9% of possible students). To gauge the effectiveness of this intervention, the Calculus 1 grades of the bailout students who took Calculus 1 the next semester were compared to the Fall 2004 grades of the students who elected to remain in Calculus 1. The mean and standard deviation of the bailout students' grades was 2.48 and 0.88 respectively, compared to 1.91 and 0.72 for the Fall 2004 non-bailout students, suggesting that the course had a positive impact on calculus performance. Koch and Herrin (2006) Though it is somewhat unfair to compare the performance of students in different semesters, and though selection bias may be a factor, this study provided motivation for the researcher to investigate Bailout Precalculus at Kansas State.

Another study similar to the current work was done by Mary Pilgrim at Colorado State as her dissertation. In this study, Pilgrim examined the effects of students switching from Calculus 1 into a course (upon poor exam 1 performance) called "Concepts for Calculus". The objective of this course is described as "...to help students gain a deeper understanding of mathematical functions and the importance of those functions as calculus tools." Pilgrim (2010). The primary difference between this thesis and Pilgrim's work is in the nature of the "bailout" course. While Concepts for Calculus (MATH 180) is similar to precalculus in that it seeks to prepare students for calculus, it is not a true precalculus course as it didn't spend a significant amount of time reviewing algebra and trigonometry concepts. This course was group based, and many of the assignments had a significant writing component in order to get the students to think of the concepts as opposed to learning procedures.

Pilgrim found that the students who completed this course had a more positive outlook on mathematics than those who stayed in Calculus 1 and failed or received a D. However, upon interviewing the students, Pilgrim learned that many of them believed that there was little connection between the two courses. Since Calculus 1 at Colorado State is a more traditional course than MATH 180, the students felt that answering open ended questions would be of little use to them in Calculus 1. Despite these feelings, students who took MATH 180 and retook Calculus 1 ended up scoring better than those who repeated Calculus 1 but declined the invitation to take MATH 180.

Chapter 3

Analysis of Calculus 1

In this chapter, the analysis of Calculus 1 using data mining and student interviews will be discussed. This study took place over three semesters at Kansas State University - Spring and Fall of 2016 and Spring of 2017. Student interviews were conducted for the Fall of 2016 and the Spring of 2017, but not for Spring of 2016.

Spring 2016

The analysis for this semester was a “pilot study”, as the researcher wished to practice using clustering algorithms on data sets of this type. The Calculus 1 class this semester had a combined total of 316 students who completed the course. In this analysis, the researcher only included students who finished the course, so the clusters created at the time of the first exam could be compared to the “more accurate” clusters formed with all the possible data at the end of the semester. Only 12 students dropped the course in this time, so the effect of dropping these students is likely insignificant. Also, any student who did not complete the first exam were excluded, since it is difficult to assign them to a group without this piece of information. It is also worth noting that Calculus 1 in the Spring has significantly fewer students overall than Calculus 1 in the Fall, as well as fewer first semester freshmen, so these results may not be applicable to Fall semesters of Calculus 1.

After the first exam, there were 301 students in Calculus 1. By this point the students had completed two homework assignments, three online homework assignments, and the exam itself. In addition, attendance data wasn't collected, so the data was only six dimensional. The researcher created a 301×6 matrix which represented each student and their scores. Then, the singular values of this matrix were computed, and are given in figure 3.4. In this

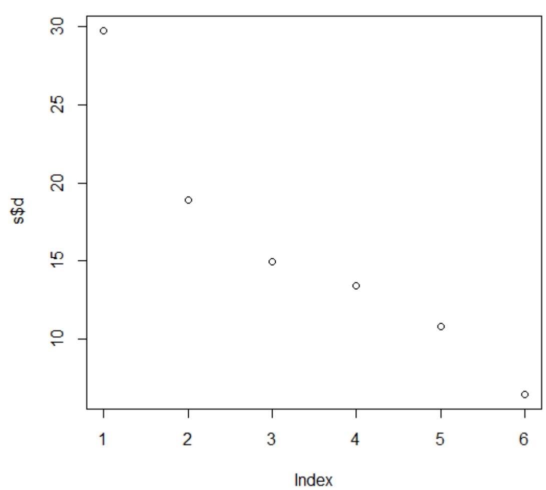

Figure 3.1: *Singular values for the Spring 2016 data.*

graph, it appears that the "elbow" of the singular values occurs at the third one, meaning this data should be truncated to three dimensions. Upon doing this, the new 301×3 matrix was run through AGNES, producing this dendrogram: In this figure, there appear to be two sensible places to cut off the AGNES process - just before and just after height 0.6. Cutting off before 0.6 produces 6 clusters, while doing so afterwards makes 4. Here is a plot of the results of PAM with 4 clusters.

Each different symbol denotes a student belonging in one of the four clusters. Note that this image is a two dimensional projection of the three dimensional data set, which accounts for the overlapping symbols. The sizes of each of the four clusters are 134, 90, 53, and 39, which are good sizes since none of them are too small (though the largest cluster is almost half of the class). When six clusters were computed, the sizes of each were 85, 84, 58, 42, 28, and 19. While there may be six legitimate "types" of students in this course, this clustering is

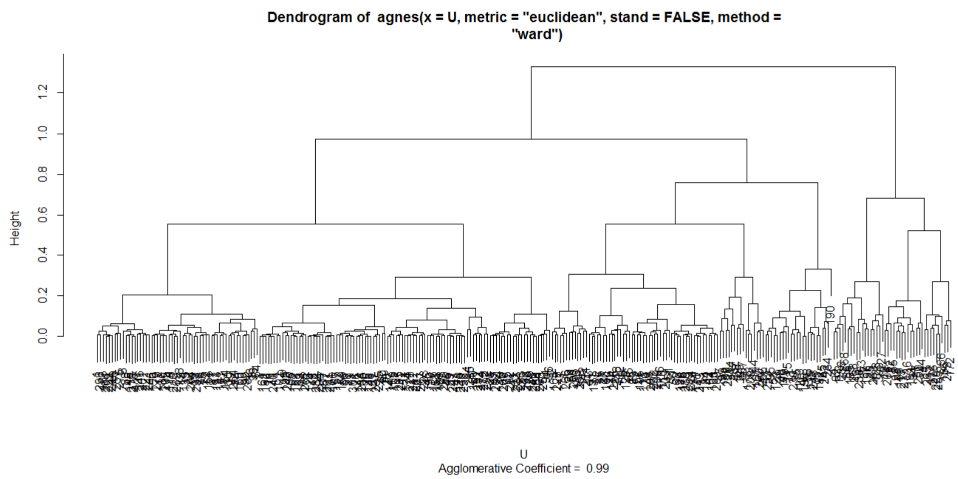

Figure 3.2: *Dendrogram for the Spring 2016 data.*

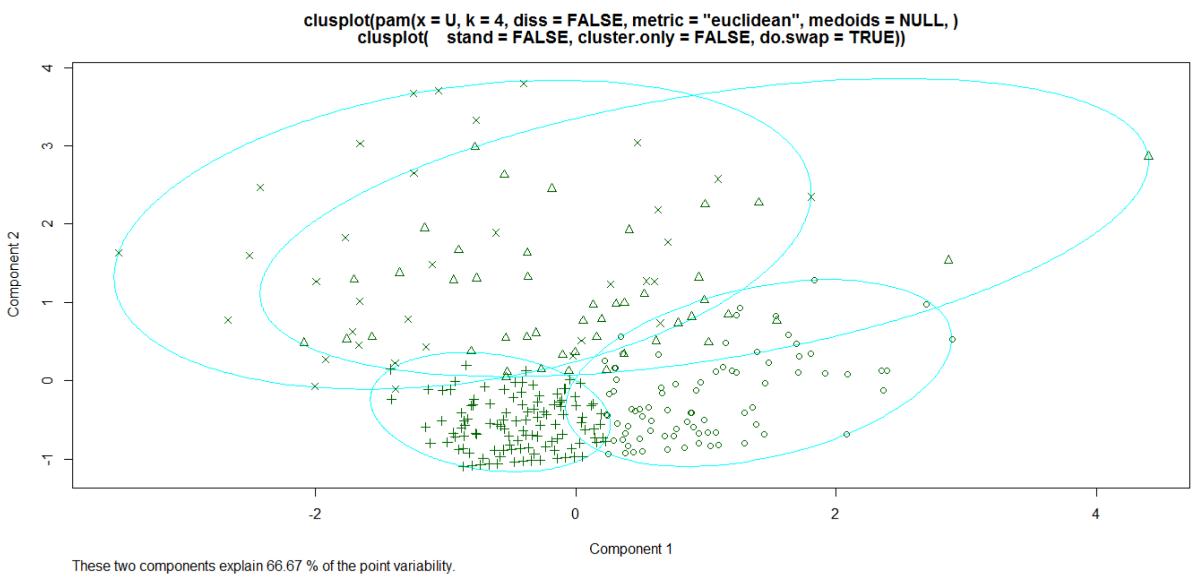

Figure 3.3: *PAM with four clusters for the Spring 2016 data.*

less useful as the two smallest clusters are too small to reliably attract students to interview for qualitative data.

To compare how consistent the clusters remain throughout the semester, this analysis was repeated after the second exam, and at the end of the course. After the second exam, each student had a total of 21 assignments (seven written, 12 online, and two exams), and at the end of the course they each had 42 grades (14 written homeworks, 4 exams, and 24 online homeworks). Below are the singular values, dendrograms, and PAM clusters of the class at each time period:

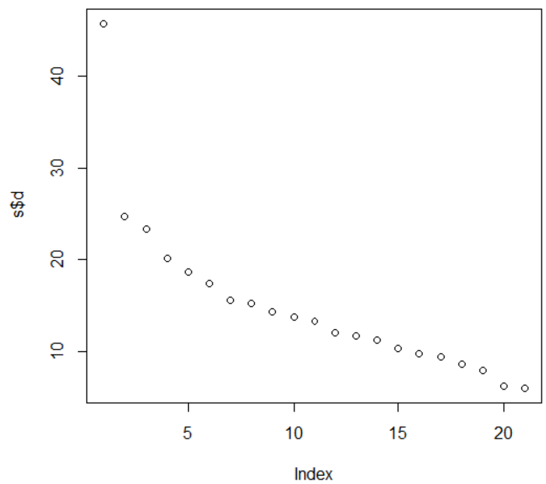

Figure 3.4: *Plot of the Spring 2016 (2 Exam) Singular Values*

In all three cases, four clusters seemed to be the best choice, given the results of AGNES. This consistency makes it easy to compare the clusterings over time. Ideally, the students in each cluster should remain mostly consistent throughout the semester, as students tend to form study and performance habits early on. Here is a table showing what percentage of students changed clusters over time, and which clusters they switched into.

Table 3.1: *Cluster Transitions Between First Exam and Final*

From/To	Group 1 (After Final)	Group 2	Group 3	Group 4
Group 1 (First Exam)	32.2%	20.1%	41.4%	5.7%
Group 2	45.7%	47.8%	4.3%	2.2%
Group 3	23.1%	18.7%	58.2%	0%
Group 4	0%	5.9%	2.9%	91.2%

The clusters were unstable both in the first exam/final case and the second exam/final case. In the first case, there were more students from group 1 that switched to group 3 than remained in group 1. Group 2 was almost as unstable, with roughly equal numbers switching to group 1 as staying in group 2. Groups 3 and 4 had a majority of their members

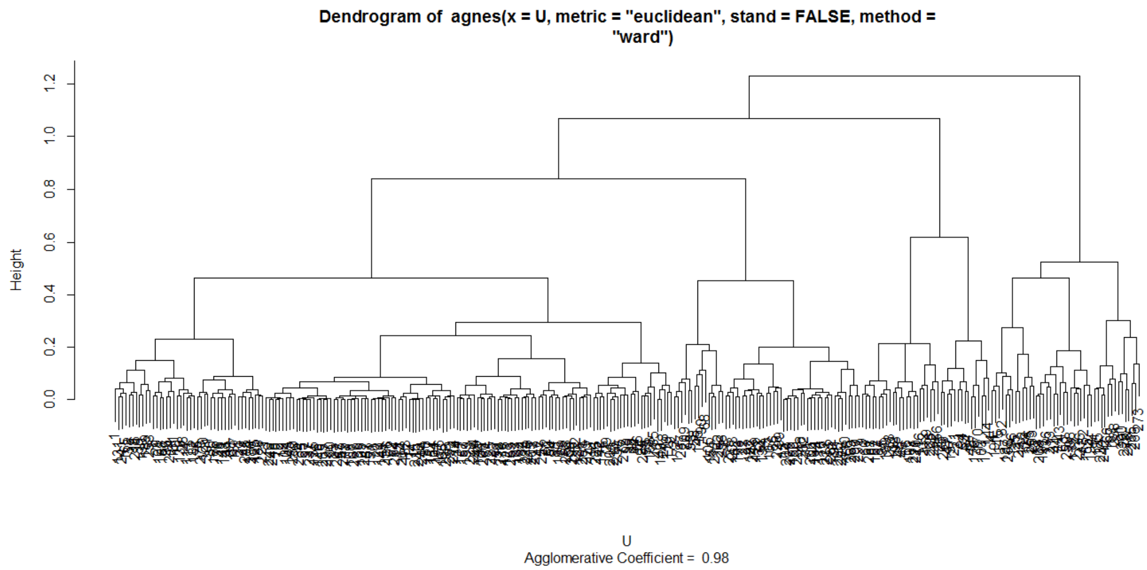

Figure 3.5: *Spring 2016 (2 Exam) Dendrogram*

Table 3.2: *Cluster Transitions Between Second Exam and Final*

From/To	Group 1 (After Final)	Group 2	Group 3	Group 4
Group 1 (First Exam)	48.8%	27.5%	13.8%	10%
Group 2	1.5%	35.8%	31.3%	31.3%
Group 3	9.5%	0%	69.8%	21.6%
Group 4	32.4%	10.8%	0%	56.8%

staying consistent. For the second exam/final case, there was a plurality of students from each group staying in their original group, though many students still switched. In the case of groups 2 and 4, the clusters were more consistent in the first exam/final case, which is surprising given that there was more data available.

This lack of consistency could be for several reasons. It is possible that the quantity and/or variety of data may not have been sufficient to create stable clusters. The lack of stability may have also been caused by the clustering algorithm. Though AGNES suggested in each case that four clusters was the best choice, it may have been a better fit for one time than another. Finally, it could be the case that many students changed their behavior throughout the semester. Though possible, this is highly unlikely, given the nature of the course and previous research Cusino (2013) suggesting that students become "set in their ways" early on in the semester.

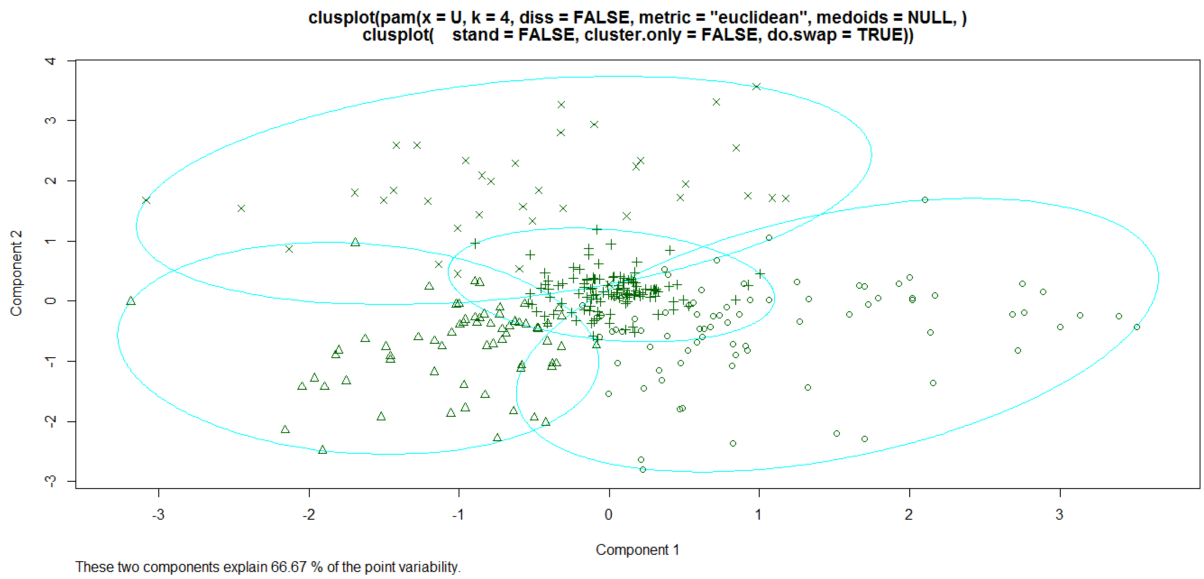

Figure 3.6: *Spring 2016 (2 Exam) PAM Clusters*

For both situations, groups 3 and 4 appeared to be the most stable. In particular, group 4 from the first exam clustering was by far the most stable, having retained 91.2% of its members after the final. Members of this group did poorly on the first exam; they received an average of 73.9% as opposed to 80.4% for all students, with a standard deviation of 14.2. Being a member of this cluster is a strong indicator of your overall performance in the course. Out of the 34 students in Cluster 4 (for the first exam), 8 received F's, 10 received D's, 8 received C's, 7 received B's, and only one student achieved an A as their final grade. This finding is supported by previous research that shows a strong relationship between performance on the first midterm exam and performance in the course overall. Unfortunately, since this group is the most stable, it suggests that students who do poorly initially are likely to continue doing so throughout the semester.

These results appear discouraging, as one would hope that the clusters would be stable over time. However, despite this instability, this pilot study shows that by using the data up to an including the first midterm, it can reliably be predicted which students are most in need of the bailout precalculus course. This was even the case when there were only six assignments used to partition the students. In the other two iterations of this study, the first exam was later in the semester, so more homework assignments and online homework

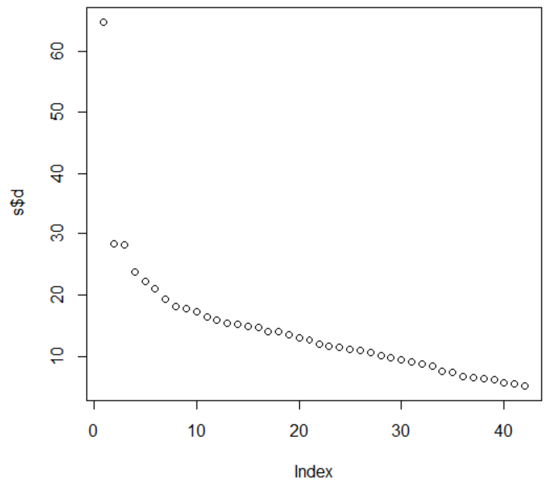

Figure 3.7: *Plot of the Spring 2016 (After Final) Singular Values*

assignments were included in the data. There was also attendance data collected, adding to the amount of variation among the students.

Fall 2016

The author conducted this study again in the Fall semester of 2016, this time with attendance data and student interviews. There were 784 students enrolled in Calculus 1 this semester. Among these, there were 449 students who both took the first exam and had their first few weeks of attendance recorded by their teaching assistants, so these students constituted the researcher's sample population. Unlike the previous semester, the first exam was held later on September 22nd, allowing for four written homeworks, six online homeworks, and ten days worth of attendance to use for data. Since attendance made up 10 of the 21 assignments and was graded in a binary fashion (1 for present, 0 for absence), the researcher believed that it may have a disproportionately high influence on the clustering. Thus, they tried clustering in several different ways - one where each day of attendance has its own score, one where the weekly attendance counts as one score, or one where all of the attendance counts as a single assignment. After examining the various plots of the singular values,

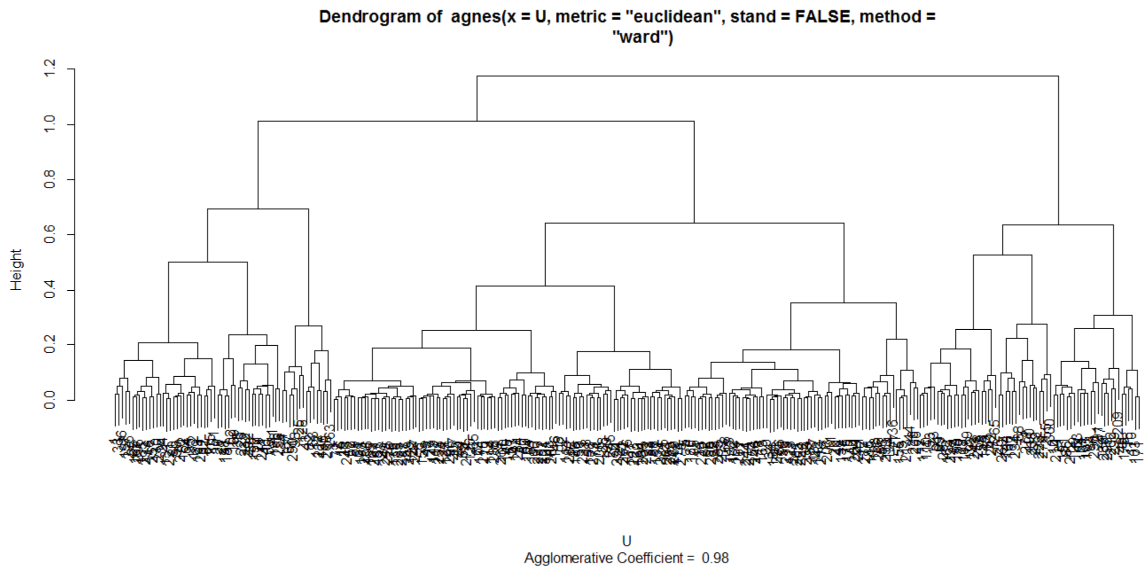

Figure 3.8: *Spring 2016 (After Final) Dendrogram*

dendrograms, and cluster sizes, the researcher decided that viewing the attendance as a single value seemed to be the best clustering. The singular values, dendrogram, and PAM results for this are included below. The interested reader can find the best (though not necessarily good) clusterings for the other two cases in the Appendix.

As before, the author first examined the singular values to determine which dimension the data should be. Typical of educational data sets, it is difficult to determine where the "elbow" of the values should be. For these values in particular, the researcher thought that it should be at either 2 singular values or 4. Since reducing the dimension of the data down to 2 would likely eliminate a significant part of the variance, the researcher chose to use 4 singular values. Then, AGNES was performed to determine the appropriate number of clusters for PAM: After examining this plot, the researcher investigated using five clusters, as this seemed to be the number suggested by AGNES. Unfortunately, after running PAM with this many clusters, two of them had less than 10% of the students, which would make it too difficult to recruit members for interviews. Four clusters was the second best option according to AGNES, so the researcher tried this many clusters in PAM instead. This time, there were no clusters with fewer than 10% of the students, and only one with around 10%, so the researcher settled on this clustering.

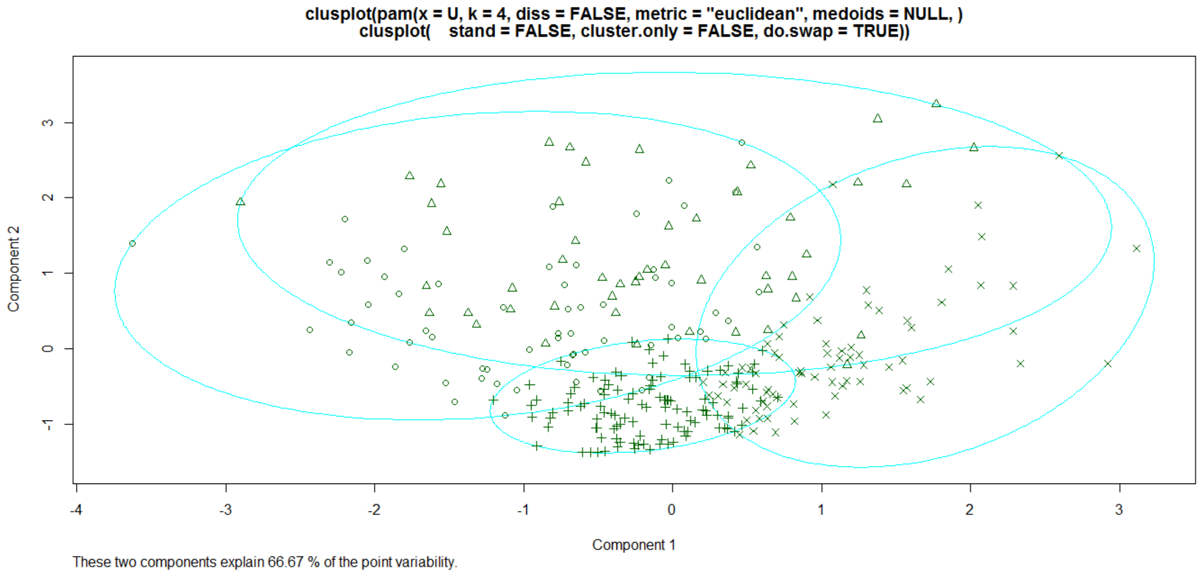

Figure 3.9: *Spring 2016 (After Final) PAM Clusters*

Now that the composition of each cluster was calculated, it was necessary to interview people from each cluster to determine the "personality" of typical members. First, the researcher computed the students who were closest to the medoid of each group and invited 20 of them in for interviews (5 for each cluster). Despite the prospect of a \$15 reward for coming in, the researcher only received several replies, none of whom were in the two smallest clusters (45 and 71 students). After this poor turnout, the researcher solicited all the students in the course to come to an interview for \$15, which convinced 18 students to participate.

Among the students interviewed, there were two from cluster 1, two from cluster 3, none from cluster 4, and the remainder from cluster 2. Even though cluster 2 was the largest (196 out of 449 students), it was still over represented, getting 77.8% of the interviews despite making up only 43.7% of the population. Since the numbers from the other clusters were so small (or 0 in the case of cluster 4), it is difficult to make assertions about those clusters "personality". However, the researcher did make some observations.

The students from Cluster 1 seemed to show a preference for recitation over lecture for learning the material. They were able to answer the conceptual and written questions well, and were doing well in the class overall. When asked what they thought math was all about,

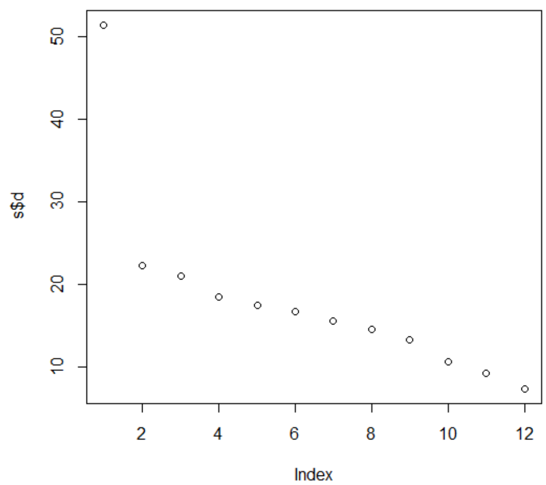

Figure 3.10: *Plot of the Fall 2016 (First Exam) Singular Values*

they mentioned critical thinking skills and solving problems. Calculus 1 was what they expected it to be, and they acknowledged that it was different from high school. The most notable feature about cluster 1 was their amount of study time - both students interviewed spend 4-6 hours a week studying for Calculus 1, which was much greater than average.

Cluster 3 was interesting in that both students interviewed claimed that math “isn’t their thing” and that they found the course difficult, but were both getting A’s in the course. They also had very practical answers for “what is math all about” and “why do we spend so much time learning about math”; math is useful for everyday life and for solving problems. These students also studied more than average, though not as much as those students in cluster 1. There was no consensus about which part of Calculus 1 was more useful - one favored lecture (due to their particular instructor) and one preferred recitation.

It was difficult to determine any information about clusters 2 and 4, since there were no interviews from cluster 2, and cluster 4 had such a high percentage of the interviews that their answers were inconsistent. Since the distribution of the sizes of the clusters were the same in almost all configurations of singular values and number of clusters, the researcher surmised that either Calculus 1 has a homogeneous student body, or there was an insufficient variety of data to distinguish many students.

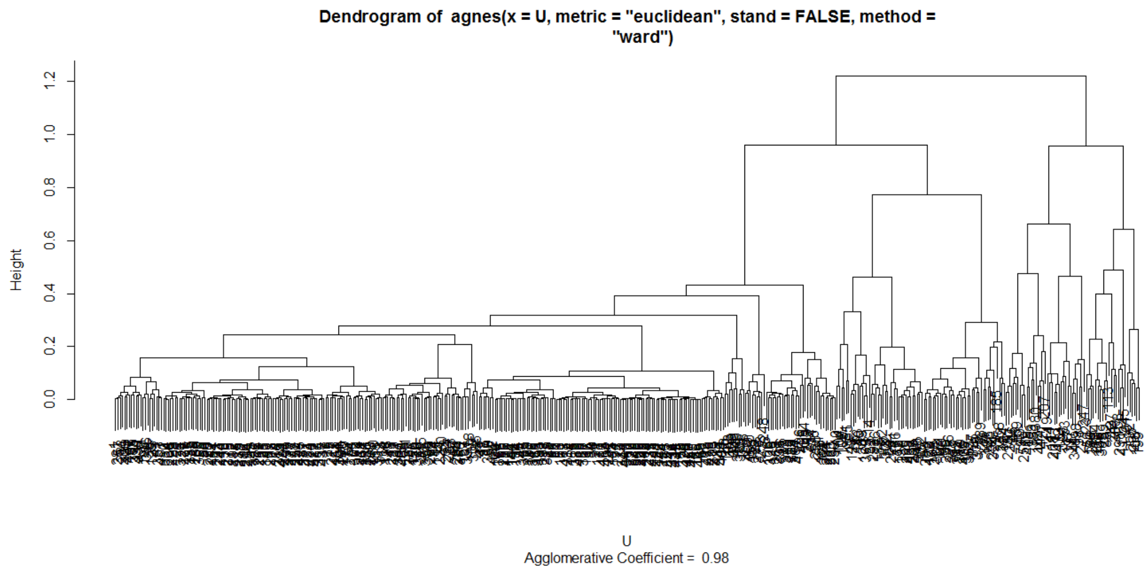

Figure 3.11: *Fall 2016 (After First Exam) Dendrogram*

Spring 2017

The following semester, the researcher performed the same study again to observe how the clusters varied between the fall and spring. Typically, students who take Calculus 1 are more likely to be students who either failed Calculus 1 in the fall and are retaking it, or are students who only need one semester of Calculus for their major. Both of these demographics are more likely to perform poorly in the course, so some differences were expected between this semester and the previous one.

This semester, there were 341 students who completed the first exam, and there were seven online homeworks, four written homeworks, and eight days of attendance before the first exam. As before, the researcher wished to mitigate the large influence attendance has on the data, so he considered the weekly and total attendance scores as well. After examining the singular values, dendrograms, and PAM clusters from various configurations, the researcher concluded that using each attendance value as an individual score with three dimensions of data and four clusters made the most coherent groups. Below are the singular values, dendrogram, and PAM clusters for this configuration. The best configurations for the other methods of attendance collection are included in the appendix.

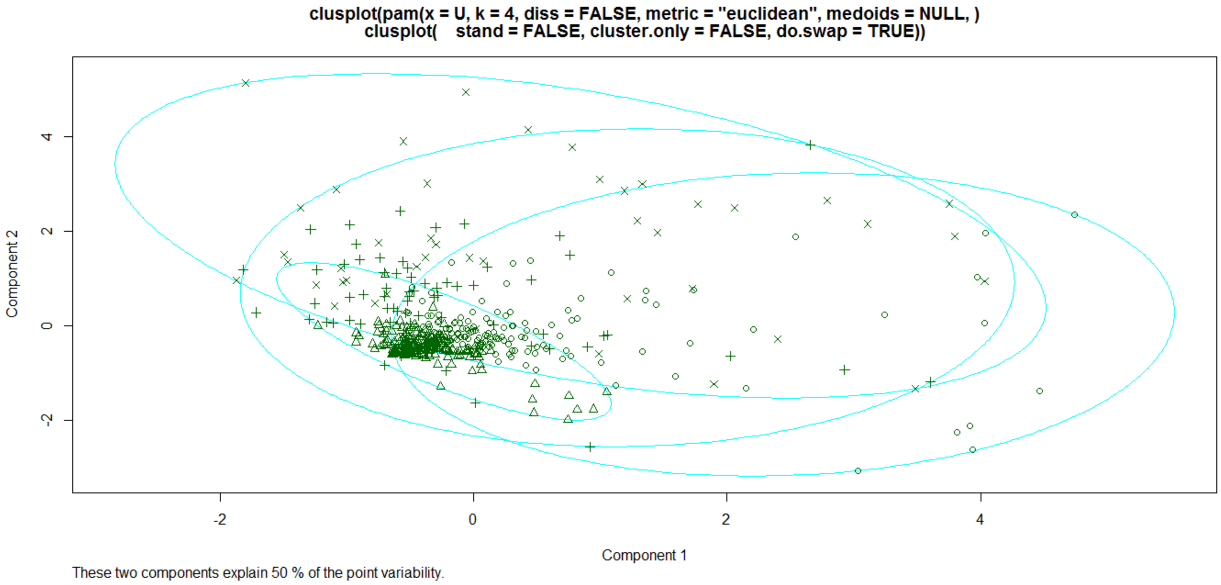

Figure 3.12: *Fall 2016 (After First Exam) PAM Clusters*

Just as it was in the previous semester, this clustering was very lopsided - about half of the students (173 out of 341) were in cluster 3, with clusters 1,2, and 4 having 68, 40, and 60 students respectively. Thus, it was difficult to find students for the smaller sections to come interview. The researcher again solicited student who were closest to the medoid of each cluster, but as before, there were not enough responses. The researcher opened the study to all students, this time through an online survey to make participation easier. This survey usually took no more than 5 minutes, and only had 10 questions (including asking the students for there ID number). A copy of the survey used is included in the appendix. There was a total of 51 responses, though only 43 of the surveys were usable as the other 8 students failed to submit their K-State ID number. There were 6 students from cluster 1, 3 from cluster 2, 28 from cluster 3, and the remaining 6 were in cluster 4.

Though the number of students in the smaller clusters (especially 2) was low, some trends were observed. Group 1 did not deviate from the rest of the class in a statistically significant way, which may be due to the low number of responses. They seemed to be more likely to be freshmen, had a stronger than average inclination towards pursuing college for knowledge, and had preparation for Calculus 2 as their primary goal for Calculus 1.

It was especially difficult to glean any insights about cluster 2, as there were only 3

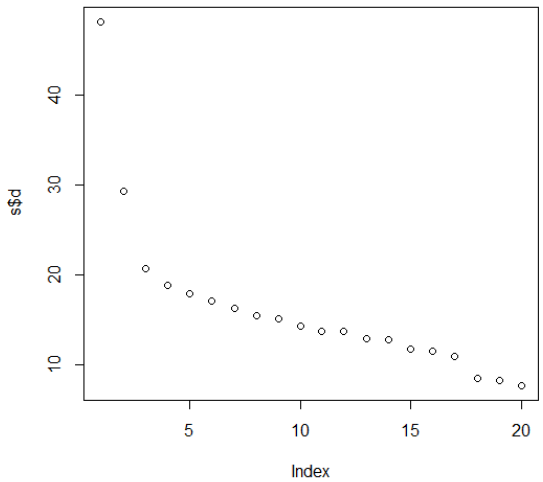

Figure 3.13: *Plot of the Spring 2017 (First Exam) Singular Values*

people who took the survey from this group. This group seemed to be the most frustrated and discouraged of the bunch - they commonly expressed that they simply wanted to get through the course, and one student wrote an essay describing their hatred towards Calculus 1. Group 3 was also difficult to analyze for the opposite reason - since they made up roughly half of the sampled population, they was no significant difference between them and randomly chosen students from the Calculus 1 class.

Cluster 4 had the greatest deviation from the rest of the course. First, five out of six of the cluster 4 students surveyed mentioned greater earning potential as a primary reason for going to college, which is over 10% more than typical. Half of them had taken Calculus in a previous semester at Kansas State, and only a third of them were Freshmen, also atypical of the population at large. Only one out of the six claimed that they were a good test taker, as opposed to roughly half of the sampled group. They all stated that seeking help was something they were likely to do if they encountered a difficult problem. Also, all but one expressed that “passing the class” or “getting an A” was what they hoped to get out of the course. Interestingly enough, the members of this cluster didn’t do much worse on the exam than the rest of the class (a 74.8 average vs a 76.3). Overall, this cluster seems to primarily be composed of students who are displeased with the course and simply want to get through

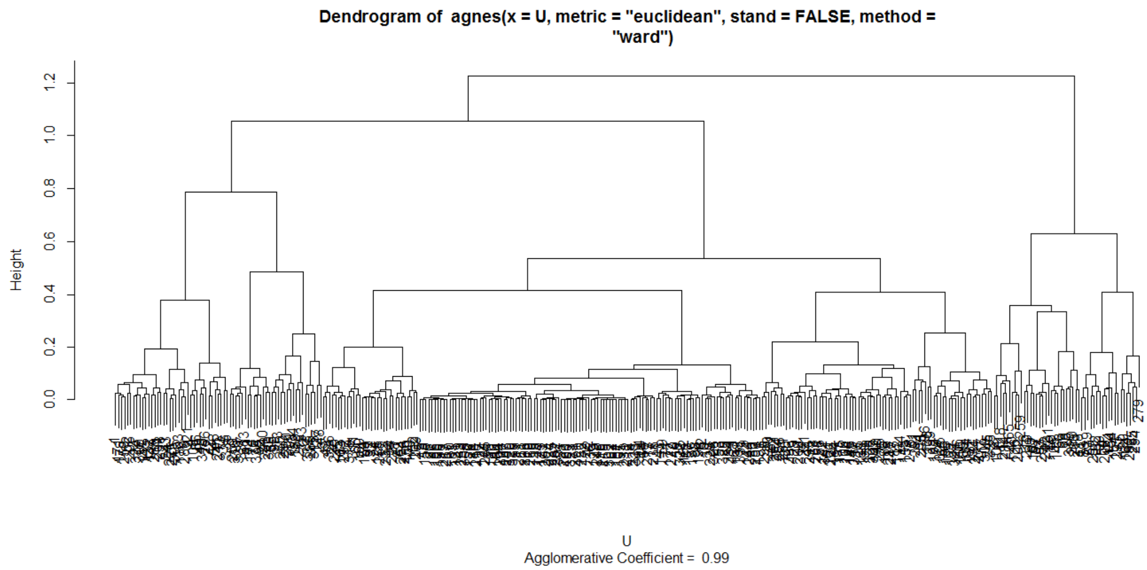

Figure 3.14: *Spring 2017 (After First Exam) Dendrogram*

so they can proceed with their major.

Conclusion

Based on the preceding information and analysis, the researcher believes that Calculus 1 has a relatively homogeneous student population. A typical Calculus 1 student is Freshman engineering major (74% of students enrolled in Calculus 1 are engineers Department (2018)) who has had some exposure to Calculus in the past. They view mathematics as a useful tool for solving problems and understanding the world (though their interest in math doesn't often extend past practical uses). They tend to view math somewhat positively, study several hours a week, prefer recitations to lecture, and use practice exams to study for the midterms. This picture is true or mostly true of the majority of students in the course, unlike College Algebra which had a more diverse population. Thus, even with data mining techniques, it remains difficult to partition Calculus 1 into groups for the purpose of differentiated instruction. This would make strategies like creating one or more alternate "styles" of Calculus lectures for different types of students difficult.

In the Fall of 2016, while the researcher was unable to find the "personality" of the

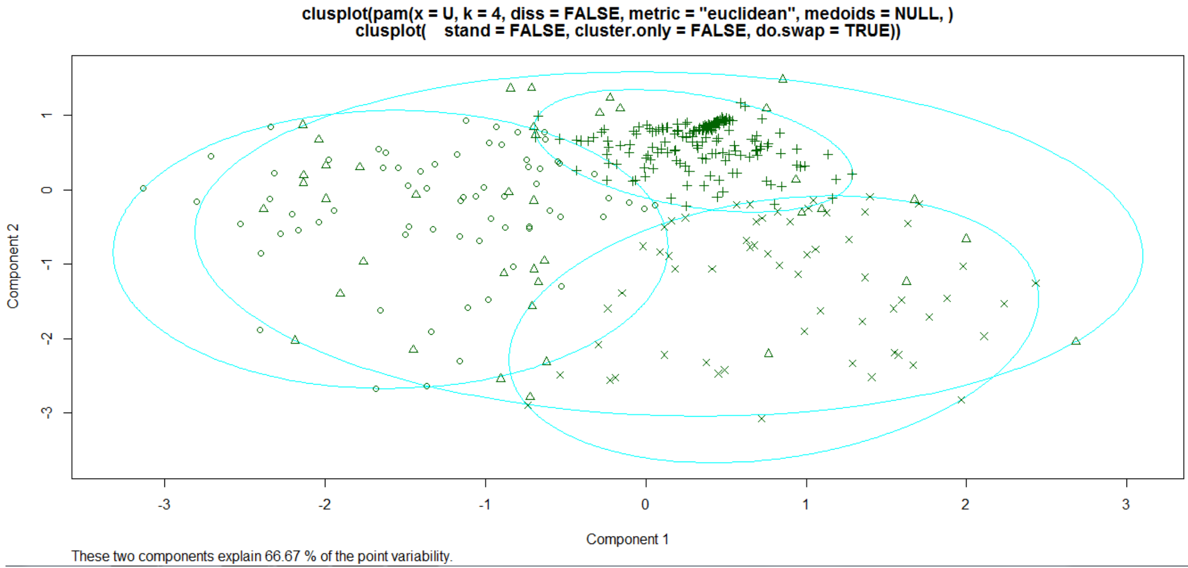

Figure 3.15: *Spring 2017 (After First Exam) PAM Clusters*

clusters that were created, they were able to find a cluster which contained a large number of struggling students. This demonstrates that it is possible to determine whether students are struggling early on in the semester. Therefore, rather than breaking the entire course into different groups, it would be more fruitful to find those students who are having difficulties and reach out to them in particular. This motivated the researcher to investigate Bailout Precalculus, and how effective it is at helping these students.

Chapter 4

Bailout Precalculus

As was mentioned in the conclusion of the previous chapter, Calculus 1 at Kansas State appears to be a relatively homogeneous class, though it is possible to detect struggling students early on in the semester. This is the focus of this chapter - what can be done to help these students? The author purposes that Bailout Precalculus is an effective tool both for retention and improving students attitudes about mathematics and their own abilities.

Bailout Precalculus Format

At Kansas State University, there is no traditional Precalculus course. If a student wants to prepare for Calculus 1, they are often advised to either take a semester of Trigonometry, or spend a year taking College Algebra, then Trigonometry (depending on the background of the student and the opinion of the advisor). Also, depending on the student's major and advisor, they may be pressured into taking Calculus 1 immediately, even if their grades and background suggests that they aren't prepared. This is done to keep students on track with their intensive major and because many scholarships for engineering students only last for four years. To remedy both this and the lack of a traditional Precalculus course, Dr. Andrew Bennett, a mathematics professor at Kansas State, created Bailout Precalculus.

When a student enrolled in Calculus 1 does poorly on their first exam (receiving a D or an

F), they are given the opportunity to switch into a newly created section of Precalculus that begins one or two weeks after the exam. Note that students can only enroll in this course via this invitation - it is not open for general enrollment. Any record of their time in Calculus is deleted, and with the help of the provost's office, students are retroactively enrolled in the new course. Unlike Calculus 1 which is four credits, Precalculus is only three, so the students are refunded a credit's worth of tuition, even though the transition typically occurs after the 100% refund date. Course times are created to match the Calculus 1 lecture time so as to not create any scheduling issues. The class meets twice a week for 50 minutes, during which students listen to lectures, work on problems, and ask questions on the homework. When the researcher asked questions about the Precalculus lecture during interviews with students who have taken the course, many students insisted that it was really a recitation, not a lecture. This reflects the course's smaller (only 30 students per section), more intimate feel, which is exactly what many students were missing when enrolled in Calculus 1. It also helps that the instructors for the course have consistently been some of the higher performing teachers at Kansas State's math department (according to teaching evaluations).

The mathematical content of the course is typical for a Precalculus class - some algebra and trigonometry review along with a dip into the early stages of Calculus 1. What makes this course unique is its emphasis on manipulating expressions into more convenient forms. This is an extremely important skill for calculus, one that is often neglected in courses like College Algebra and Trigonometry. Shown below are some examples of questions students may be asked in Precalculus:

$$\frac{3t}{2} = 2t[\qquad]$$

$$\frac{\sin(3t)}{t^2+t} = \frac{\sin(3t)}{3t}\frac{[\qquad]}{[\qquad]}$$

Write $\sin^3(t)$ in the form $\sin(t)f(\cos(t))$.

Write $\cos^2(2t)$ in the form $g(\cos(4t))$.

The first example is something that may be required when performing a u-substitution, and knowing how to do the following problem could be helpful in evaluating a difficult limit. The last two examples have students put trigonometric expressions into more convenient forms for integration. This course emphasizes to students that while knowing algebra and trigonometry in general is useful, being able to manipulate these particular types of expressions is immensely valuable.

The final few weeks of the course are spent relearning the calculus that the students encountered their first month in Calculus 1. While having the obvious effect of preparing them to encounter the material again in Calculus, it also has an impact on the students' mathematical self-esteem. If the students can master the material that they previously had difficulty with, it can help them get into a growth mindset. They can realize that now that their prerequisite skills are solidified, they are ready to take on the challenge of Calculus 1 again, if they so choose.

Assessment in the course is standard for a lower level college math course - weekly homework, online problems, a midterm and a final are used to calculate the students' grades. Students do well in this course, with 75% getting a C or better (this percentage increases to 88.9% when students who didn't participate in the course are removed from consideration). Given that all of the students who enrolled in the course were failing or doing poorly in Calculus 1, this constitutes a significant improvement to their GPA.

Impact of Bailout Precalculus

What sort of impact does this unconventional course have? The most easily observed effect is that of retention. The majority of students who will leave the university before graduating do so before their Sophomore year OPA (2018), commonly due to difficult courses, not enough personal attention, or monetary problems. As four-year graduation rates are a common metric for the quality of a school, colleges and universities are always trying to retain students. One of the functions of Bailout Precalculus is that it serves as a way to get students to remain in school, despite having an initial failure.

In the Fall of 2013, when the course was first offered, 39 students enrolled in Bailout

Precalculus. Note that enrollment was smaller than in the following semesters since there was only one instructor available to teach the course. Out of these 39 students, 85% were still enrolled at Kansas State a year later. On the other hand, among the at-risk students who remained in Calculus 1 despite doing poorly on their exam, only 41% were enrolled the following Fall. However, these results are susceptible to selection bias, as failing the first midterm is a strong indicator of performing badly in the course overall. To account for this, The retention rates of the students who failed Calculus 1 in 2012 were compared to the students who failed the following year combined with the Bailout students. 51% of the 2012 students remained after a year, compared to 55% of students in the Bail/Fail group from 2013. This supports the assertion that the Bailout Course does indeed improve retention, at least in the university as a whole.

How does Bailout Precalculus affect retention in the students' majors, or at least within STEM? Of the 44 students who took Bailout Precalculus in 2014, 18 retook and passed Calculus 1 within the following year. Their average score in the course was a 2.23 out of 4, and 15 of the 18 achieved the necessary C or better to continue on in the calculus sequence. As for the 80 students who failed Calculus 1, 18 retook Calculus 1 and 10 passed with a C or better. Again, selection bias may be at work, so the researcher compared this with the Fall of 2012. That semester, 94 students failed Calculus 1, 33 retook the class within the following year, and 25 got a C or better.

Table 4.1: *2012 Fail vs. 2014 Bail + Fail*

	2012	2014
Fail or Bail + Fail	94	124
Retook and Failed	1	3
Retook and Passed	32	37
C or Better	25	25

Thus, the percentage of students in 2012 who failed but got a C or higher within one year is 27%, whereas the percentage of students who either bailed or failed in 2014 who did the same was only 20%. This suggests that Bailout Precalculus doesn't increase the number of students who continue on in their mathematical training. However, note that a higher percentage of students in 2012 (35%) decided to give calculus another try, as opposed to only 32% of the Bail+Fail group. This can be attributed to the fact that many students decide to change their major to something that doesn't require Calculus 1 upon switching into Precalculus, so they never intend to retake the course at all. Despite this, 78% of the 2012 students who retook the course got a C or better, which is higher than the 68% for the 2014 students. Overall, it appears that Precalculus does not have a positive impact on retention in STEM. However, this is not different from how college level Precalculus courses typically run. In a study conducted by Sonnert and Sadler in 2013, taking a precalculus course in college has no effect on subsequent performance in Calculus Sonnert and Sadler (2013).

The researcher questioned what made this course have a positive effect on retention as a whole, while not increasing the retention in STEM. To address this issue, the researcher conducted in-person interviews with students from the Fall of 2016 and the Fall of 2017 (Precalculus is not offered in the Spring). These interviews were conducted in November of 2017, shortly before Thanksgiving break, to give the current Precalculus students time to form an opinion of the course. The students were given $15 upon completing the interview. The briefing, debriefing, informed consent, and interivew procedure forms are included in the Appendix.

The researcher went into both sections of the 2017 course and solicited students to be interviewed. Unlike the interviews conducted with the Calculus 1 students, these interviews had a higher response rate, with 20% (12 of 60) students volunteering from the 2017 semester, and 6 students from 2016. The interviews consisted of 15 questions created to assess the students' opinions about their time in Calculus, Precalculus, and mathematics in general.

The first question in each interview was "What is math all about?". The researcher included this question to determine what Precalculus students think about math in general and which aspects of it are most important to them. As expected, 13 out of the 18 students said that it was primarily "used in calculations and problems", with the 2nd most common answers being "understand the world better" and "numbers". Only one student came close to describing math independently of its practical uses, stating that it is a language.

The responses to this question suggest that typical Precalculus students view math almost exclusively as a problem solving tool.

Following this question was “What was your opinion of math before taking Calculus 1?”. A majority of the people interviewed (11 out of 18) interpreted this question as “how difficult did you find math”, with 5 thinking it was difficult, 3 ok, and 3 easy. The other responses were not particularly fleshed out, as they typically responded by saying they liked it or they didn’t like it. Four of the 18 mentioned that it got more difficult in college, and two admitted that math was only a means to an end for them.

When asked about how their opinion changed after switching into Precalculus, half the students claimed that things got better for them, with 3 saying their opinion stayed the same. Participants mentioned various other changes, such as how math got more interesting, how they liked the detailed explanations given in the course, how they liked being able to practice problems, and how the class gave them more confidence. These responses support the claim that this course boosts retention by giving students more personalized attention, slowing down to allow for more detailed explanations of complex topics, and by instilling more confidence in students.

After this, the researcher asked the students what their major was before Precalculus and whether or not it changed upon bailing out. All of the students being interviewed began in a major that required at least Calculus 1. Half of the students decided to change their major, and among these 9, 8 transitioned to a major that required only Business Calculus or no math at all. Six switched to a Business Calc major, and 2 switched to non-mathematical majors. The student who switched but still needed to take Calculus changed their major from the general engineering track to secondary education in mathematics, having been inspired to do so from taking Precalculus. Thus, it appears that many students decide that their STEM major isn’t for them before retrying Calculus 1.

As described earlier in this work, the vast majority of students enrolled in Calculus 1 in a Fall semester are freshmen. However, the Bailout Precalculus class is more varied. Only half of the students being interviewed were Freshmen, five were sophomores, and three were juniors. This suggests that older students may have postponed their mathematics requirement for later, which is possible for certain STEM majors, and may have had their mathematical skills deteriorate over time.

When asked about their experiences in calculus before the semester they enrolled in Precalculus, only four of the eighteen students said they had taken calculus before. Note that this question was worded in a way to include high school calculus, implying that at most 22% of these students have had calculus before college. This a drastic difference from the Calculus 1 population as a whole, where 67% of students have been exposed to calculus prior to university enrollment Bressoud et al. (2015).

This lack of previous calculus experience explains some of the answers to the next question, "Describe your experience with Calculus 1 before Precalculus". The most common response was that the course went to fast, and that they didn't feel prepared for the class. There was a significant overlap between these responses and not having previously taken calculus. Futhermore, the interviewees also mentioned that they were intimidated by the other students in the class, particularly those who had previous calculus experience. This was among other responses such as complaints about not being able to ask questions in class, how they didn't like how the course was run, how the concepts were difficult to understand, and how calculus wasn't as straightforward as previous mathematics. These students appeared to be discouraged with their experience in Calculus 1, and were just as frustrated with the organization/format of the course as they were with the mathematics.

As to why students decided to bail out, the majority found the grade forgiveness part of the course attractive, as well as the opportunity to improve their skills in calculus and mathematics in general. Three students mentioned how they were overwhelmed with calculus, while others cited that they didn't need it for their new major or that they wanted to focus on other courses. Only one student said that their advisor suggested the switch. This suggested that this course's existence was not well known throughout the university. In fact, several students that the researcher spoke with mentioned how other departments don't understand what the course is and won't accept it for mathematics requirements, since it is listed online as "MATH 199 - Topics in Mathematics".

Overall, the interviewees gave positive responses to the question "Describe your experiences in Precalculus". Eleven said that they liked it and that it served as a good review, and three cited their positive experiences with their instructor. Two students said they enjoyed learning why things are true vs. simple memorization. There were two minor negative comments - one said that they would have liked a book, and another student wished for a class that was "in between Calculus 1 and Precalculus", but there were no serious complaints. Additionally, when asked if the course improved their math skills, all of them said yes, and some (6) elaborated to say that they left the course with a better understanding of Calculus.

When asked to elaborate on how the format of the course affected their opinion, 10 of the students said that they liked the "more personal" environment in the class, and four of them said that it reminded them of high school. There were other positive responses, such as how the course went at a good speed, how they got to do more problems in class, how they enjoyed only meeting twice a week, and that they got to ask questions, unlike in large lecture. Only one student mentioned that they missed the Calculus 1 format. Upon being asked if they missed recitation from Calculus 1, all but two students said no. They elaborated by saying that typical classes in Precalculus are more similar to recitations than lectures, with two students calling the classes recitations without prompting from the researcher. These responses support the assertion that struggling students desire a smaller class and that having this setting in Precalculus enables them to succeed.

The following two questions were only asked of those students who had taken the course in 2016 or earlier. The first was "What grade did you receive in the course?". 5 of the 6 got a B, with the other receiving and A (there are no +'s or -'s at Kansas State University). They were also asked if they took any math since Precalculus. Three attempted Calculus 1 again, one took Business Calc, one was in Calculus II, and one hadn't taken any more math. Though the numbers are small, this shows that students who take Precalculus can indeed have success in further mathematics courses.

Finally, the interviewees were asked to rate, on a scale from 1 to 5 (5 being greatest) how impactful Bailout Precalculus was on their college career. For this question, there was a divide among the 2016 and 2017 students. The average score for the six 2016 students was a 3.83, while the average score for the other 12 2017 students was only a 2.92. This suggests that while students don't perceive the course as being a "game changer" right away, over time they realize that the course played a significant role in shaping the direction of their college career.

Conclusion

Bailout Precalculus has many desirable effects on students. It increases the likelihood that they remain in the university. Students gain a more positive outlook both on mathematics itself and their own abilities in mathematics. In particular, they no longer feel behind when compared to their peers who have seen calculus previously. Finally, the format of the course helps students feel more comfortable in the course, affording them the opportunity to focus learning mathematics rather than worrying about being lost in the crowd or being without access to instructor assistance.

Despite having these benefits, Bailout Precalculus has some weaknesses as a course. The class did not have a significant impact on the C-or-better rate of the Calculus 1 + Bailout Precalculus population, though this is not an unexpected result for a college Precalculus course. Additionally, Bailout Precalculus remains an obscure class at Kansas State, which often prevents students from using it to satisfy mathematics requirements. Despite these drawbacks, students have a positive opinion of the course and its impact, and their appreciation for the course grows over time.

Chapter 5

Conclusion

With the content of the previous two chapters, was the researcher able to find answers to their research questions?

Question 1

Can data mining break the Calculus 1 large lecture course into different groups to allow the lecturer to address each type of student? If so, what would be the nature of these groups?

Data mining yielded few interesting results about Calculus 1, despite it being successful for College Algebra. There are several reasons why this could be the case. With Manspeaker's work, there were significantly more data points before the first exam. Also, Manspeaker had access to scores on individual exam questions, which could have made the data sets in this work more varied and rich. It is also likely that Calculus 1 is simply too homogeneous to be partitioned in this manner. From this study, the researcher observed that the majority of Calculus students are engineering or science majors who appreciate the utility of math, but aren't particularly passionate about math in its own right. They commonly prefer recitation and problem solving to large lectures, and are likely to ask their friends for help and work together on assignments. They have a decent procedural understanding of Calculus, but have room for improvement in conceptual understanding. Many instructors and professors "intuitively" know this after having taught Calculus, and the interviews with students and the homogeneity of the clusters support this belief.

Question 2

For Calculus 1, is it possible to quickly identify students who are in need of extra assistance?

The researcher's data mining produced positive results for this research question. From the Spring of 2016, the researcher was able to determine which students were at risk using clustering algorithms, allowing for these students to be identified and assisted early on. This could take the form of extra help or tutoring, or perhaps an offer to switch into Bailout Precalculus. However, the majority of students identified in this way could have also been identified by looking for poor exam scores, but by examining the scores from the end of the semester, there were students who did poorly on the first exam but passed the course, so this method may still have utility.

Is Bailout Precalculus a good way of helping these students?

The information from Chapter 4 suggest that Bailout Precalculus is indeed a good way of helping students. Students in Bailout Precalculus get the opportunity to improve their mathematical skills. Rather than taking a year long Algebra - Trig sequence, or spending an entire semester on just one of these topics, Bailout Precalculus takes students through both in 10 weeks. The old material isn't just reviewed - it's shown in a way that is useful for calculations that frequently appear in Calculus 1 and the rest of the calculus sequence. This has the side effect of increasing students' confidence in their own abilities, which is certainly a boon for those who have failed previously.

Additionally, people who take Bailout Precalculus have a higher chance of staying enrolled in the University. Though they aren't more likely to be successful in passing Calculus 1 in the future (compared with students who fail Calculus 1), this isn't due entirely to academic failure. Half of the students interviewed decided to change their major to something that didn't require Calculus, so they didn't pass Calculus in the future because they never took it again.

Question 3

If Bailout Precalculus is effective at changing the trajectory of students' mathematical and educational progress, why is it effective?

Intervening in this way also helps students gain confidence in their mathematical skills and their ability as students. Since the last half of the course is spent reviewing the material that the students previously failed to master, they get a second chance at success. Given the grades in the course and the student interviews, the vast majority of Precalculus students are indeed successful the second time around. In interviews, they were appreciative of this opportunity, and felt it was important.

This course also helps students have extra time to transition from high school to college style classes. As mentioned previously, many students had not only never taken calculus before, but they were also not used to large lecture courses. Taking this course gave students the opportunity to learn college level material in an environment they were more familiar with, reducing the stress that accompanies the college transition. Giving students this "half high school, half college" experience sets them up for success not just in math but in any college course.

Further evidence of this is the difference in ratings of the course between current and past students. Students currently in the course seem to perceive it as "just another math course", and while they seem to be grateful for the opportunity to erase their bad Calculus 1 grades, they still aren't optimistic. This is in contrast to students who had taken the course a year or more in the past who gave the course a higher score, as they have had the chance to see how the confidence and study skills they gained can be used for any course.

Future Work

There are several different avenues for continuing the work described in this dissertation. As stated in Manspeaker's thesis, the data mining techniques used for College Algebra (or in our case, Calculus 1) are not unique to this particular course. Theoretically, any large lecture class can be analyzed in this way. However, any future researcher should take care to obtain robust data, in particular individual exam scores. In fact, it may not be amiss to repeat the current study with this data to see if Calculus 1 students aren't as homogeneous as they appear.

Another expansion of this work could be to see how clusters change as students advance along the Calculus sequence. By computing clusters again with Calculus 2 and 3 (and possibly Differential Equations), one could see if students change their behaviors between semesters. It may also be possible to tell in advance which types of students (beyond grades and majors) will continue or stop along the sequence. Again, one would be advised to obtain as much data as possible to insure useful partitions.

As for Bailout Precalculus, there may be other courses besides Calculus 1 where students could benefit from intervention. For example, if a student isn't prepared for College Algebra and fails the first exam, they could switch into a newly created section of Intermediate Algebra (a remedial math course taught at Kansas State University). Whether or not such an intervention is effective could be the subject of future research. However, some courses are not well suited for intervention. If a student is having difficulty with Calculus II or III, it may be better for them to go back to a lower level course like Calculus 1, though many students or advisors wouldn't be pleased with that option.

This study was conducted solely at Kansas State University, a large, Midwestern research university. If this work is repeated at a different institution, even one that is similar to Kansas State, the results may be different. It is possible that more viable clusters for Calculus 1 could be found elsewhere. Also, schools with different student demographics than Kansas State may have different experiences with an intervention course like Bailout Precalculus. What these differences are may be woth investigating.

Finally, the questions and assignments used in Bailout Precalculus were somewhat unique, in that they focused on manipulating expressions as opposed to solving equations. Future study of the efficacy of these questions in building student understanding could be useful. If these questions and activities are effective, they could be introduced into other courses like Calculus 1, or courses that are geared towards preparing students for Calculus, such as

College Algebra or Trigonometry.

Implications

During the interviews with Precalculus students, the researcher was told several times by the interviewees that they wished that Precalculus was a regular course offering at Kansas State University. At the time of writing, Kansas State University doesn't have a traditional Precalculus course - students are frequently advised to take Trigonometry or the year long College Algebra - Trigonometry combination. If there was a regular Precalculus course offered, students wouldn't have to have the experience of failing an exam (though this may be a crucial ingredient in changing student behavior).

Also, while working on this study, the researcher discovered that this course was relatively unknown throughout the university. Professors and advisors who are unaware of the course may not be able to offer good advice to students who are considering taking it. It may also not count for math credit in certain majors where it would be appropriate for Precalculus to satisfy their mathematics requirements. Finally, since the course and its benefits aren't widely known among the student body, it may be difficult to recruit students to bailout (currently, only 40% of students asked to bail out do so).

Through data mining, differentiated instruction, intervention, or other methods, the researcher hopes that students who are struggling can be helped, despite large enrollment in the modern university.

Bibliography

D. A. Bligh. *What's the Use of Lectures?* Intellect Books, 1998.

D. Cusino. *Using Bayesian Learning to Classify College Algebra Students By Understanding in Real-Time*. PhD thesis, Kansas State University, 2013.

R. Manspeaker. *Using Data Mining to Differentiate Instruction in College Algebra*. PhD thesis, Kansas State University, 2011.

D. Bressoud, V. Mesa, and C. Rasmussen. Insights and recommendations from the maa national study of college calculus, 2015. URL https://www.maa.org/sites/default/files/pdf/cspcc/InsightsandRecommendations.pdf.

Academic year tuition and required fees. Kansas State University, Manhattan, KS, 2011. URL http://www.k-state.edu/pa/statinfo/reports/tuition/ksutuition.pdf.

K. Vasquez. Learning styles as self-fulfilling prophecies. In *Getting Culture: Incorporating Diversity Across the Curriculum*. Stylus Publishing, LLC, 2009.

P. N. Tan, M. Steinbach, A. Karpatne, and V. Kumar. *Introduction to Data Mining*. Pearson, 2005.

M. Lovell. Data mining. *The Review of Economics and Statistics*, 65:1–12, 1983.

Edm 2008. International Educational Data Mining Society webpage, 2008. URL http://educationaldatamining.org/.

C. Romero and S. Ventura. Educational data mining: A review of the state-of-the-art. *IEEE Transaction on Systems, Man , and Cybernetics, Part C: Applications and Reviews*, 40: 601–618, 2010.

L. Kaufman and P. J. Rousseeuw. *Finding Groups in Data: An Introduction to Cluster Analysis*. Wiley, 1990.

E. Seymour and N. M. Hewitt. *Talking About Leaving*. 1997.

J. G. Rosenstein and A. Ahluwalia. Putting brakes on the rush to ap calculus. *The Role of Calculus in the Transition From High School to College Mathematics*, 1:27–40, 2016.

D. Koch and G. D. Herrin. Invervention strategy for improving success rates in calculus, 2006. URL https://peer.asee.org/intervention-strategy-for-improving-success-rates-in-calculus.pdf.

M. E. Pilgrim. *A Concepts for Calculus Intervention: Measuring Student Attitudes Toward Mathematics and Achievement in Calculus*. PhD thesis, Colorado State University, 2010.

Kansas State University Mathematics Department, 2018.

Big twelve longitudnal retention survey. Kansas State University, Manhattan, KS, 2018. URL http://www.k-state.edu/pa/retention/totals.pdf.

G. Sonnert and P. M. Sadler. The impact of taking a college precalculus course on students' college calculus performance. *International Journal of Mathematical Education in Science and Technology*, 2013.

Appendix A

Interview Forms, Questions, and Results

This appendix contains the informed consent forms, briefing, debriefing, and interview procedure forms for each semester interviews were performed, as well as student responses. Items are listed in order of semester. The forms used for the Fall of 2016 and the Spring of 2017 were identical. Note that only the Fall of 2017 has a debriefing form, as this was the only time it was required by the IRB.

Fall 2016

Consent Form - Cluster Analysis and Differentiated Instruction in Calculus I

April 7, 2016

You are being asked to take part in a research study on the different categories of students in Calculus 1. We are asking you to take part because you are currently enrolled in Calculus 1. Please read this form carefully and ask any questions you may have before agreeing to take part in the study.

What the study is about: This study is about using data mining techniques (cluster analysis in particular) to determine the different characteristics of students in Calculus I.

What we will ask you to do: If you are willing, we will conduct an interview with you. We will ask you questions on your opinions of the course and mathematics in general. We will also ask you to solve several computational and conceptual Calculus 1 problems. The entire interview should take no longer than 30 minutes. With your permission, we would also like to tape-record the interview.

Risks and benefits: We do not anticipate any risks to you participating in this study other than those encountered in day-to-day life. There are two benefits - the compensation (described below), and the results of the study, when it is complete.

Compensation: Upon completion of the interview, you will receive $15 cash. If you decide to terminate the interview before its completion, you will not receive the compensation. Also, you will be informed of the results of this research project via email when it is completed.

Your answers will be confidential. The records of this study will be kept private. In any sort of report we make public we will not include any information that will make it possible to identify you. Research records will be kept in a locked file; only the researchers will have access to the records. If we tape-record the interview, we will destroy the tape after it has been transcribed.

Taking part is voluntary. Taking part in this study is completely voluntary. You may skip any questions that you do not want to answer. If you decide not to take part or to skip some of the questions, it will not affect your current or future relationship with Kansas State University, or your grade in Calculus I. If you decide to take part, you are free to withdraw at any time.

If you have questions: The researchers conducting this study are Ian Manly and Prof. Andrew Bennett. Please ask any questions you have now. If you have a question later, you may contact Ian Manly at imanly62@math.ksu.edu or at (315) 935-5099. You can reach Prof. Bennett at bennett@math.ksu.edu or (785) 532-0562. If you have any questions or concerns regarding your rights as a subject in this study, you may contact the Institutional Review Board (IRB) at (785) 532-3224 or access their website at https://www.k-state.edu/comply/irb/.

You will be given a copy of this form to keep for your records.

Statement of Consent: I have read the above information, and have received answers to any questions I asked. I consent to take part in the study.

Your Signature: ______________________ Date: __________

Your Name (printed): ______________________________

In addition to agreeing to participate, I also consent to having the interview tape-recorded.

Your Signature: ______________________ Date: __________

Signature of person obtaining consent: ______________________

Date: __________

Printed name of person obtaining consent: ______________________

Date: __________

This consent form will be kept by the researcher for at least five years beyond the end of the study.

Interview Procedures

April 7, 2016

1. Prepare for the interview at least 5 minutes before the scheduled time. Unlock the conference room (Reta has the key) and leave the door open. Set out the IC Recorder, two copies of the Informed Consent form and a pad of paper for students to write or draw on as needed when they answer the questions.

2. When the student arrives, introduce yourself and welcome the student by name. Close the door to the conference room. Ask for permission to record the interview. If permission is granted, start the recorder.

3. Explain the purpose of the interview:

   We are interviewing students in Calculus 1 to better describe the characteristics of students enrolled in the class. This is prompted by a desire to understand how different students react to certain aspects of the course, how they set about learning the material, and their level of conceptual understanding. The general goal is to use this information to improve teaching and assessment. This interview should take approximately 20-45 minutes. Your participation is completely voluntary and your grade will not be affected by your answers in this interview. You will receive $10 for your time for participating in this interview and you may also benefit by improvements in instruction in mathematics and by having a chance to go over the most recent exam an instructor. In the event we include any of your comments in a discussion or publication about our findings, your privacy will be maintained by the use of a pseudonym. We have two copies of an Informed Consent Form for you to sign, one for our records and one for you to keep.

4. Have them read and sign the form. If they decline to sign the form, thank them for their time and terminate the interview. Otherwise sign and date the form as witness and then proceed to the questions below.

5. **Background/Attitude Questions.** Stay aware of the time and try not to let this section exceed 20 minutes so you have time for the rest of the material. In the (unusual) event that a student wants to spend more than 20 minutes on this, explain politely that you need to get to some additional questions and promise them they will have a chance to make more comments at the end.

   (a) *What do you think mathematics is all about? Is it important? Why do we spend so much time learning about math?*

   (b) *Describe your feelings towards mathematics at the beginning of the semester as you entered into this course. Have they changed at all over the last few months?* (Follow up questions: How?, Why/ Why not?, etc.)

   (c) *Describe your experience with Calculus 1 so far this semester.*

      i. *Have you enjoyed the class?*

      ii. *Are you doing well?*

      iii. *Is it what you expected?*

   (d) *How much time outside of class do you normally spend each week on Calculus 1 related work?* Ask them to specify which activities make up this time.

   (e) *If you get stuck on a problem or have trouble understanding a concept, what do you do?* Ask them to explain why they choose to seek help or not.

   (f) *How do you usually study for math quizzes or tests? Did you study differently for different quizzes or tests in this course? Explain.*

   (g) *How do you do the online homework in this course? Do you enter in incorrect answers in order to see a worked problem, or do you try and solve the problem on your own first?*

   (h) *What are your future career goals?*

   (i) *Which topics in Calculus 1 do you feel are important to know for your future? Explain why/why not.*

   (j) *What aspect of the calculus class (lecture, recitation, written homework, online homework) have you found most helpful?* Ask them to explain why this has been helpful.

   (k) *What aspect of class (lecture, recitation, written homework, online homework) have you found least helpful?* Ask them to explain what the problems with this aspect of the class are.

(l) *What suggestions do you have for improving the course?*

6. **Conceptual Questions** This section should take 5-10 minutes.

   Now I want to ask you a few questions about some basic concepts in calculus. After saying this, present the student with a sheet with several conceptual calculus questions.

   As you are solving these problems, please explain your reasoning for each step of the solution. Then the student should begin solving the problems on the sheet.

7. **Problem Solving** This section should take 15 minutes or less.

   This concludes our interview. (Thank the student for participating) *Before you leave, Id like to know if there were any questions I should have asked you, but I missed.* Ask follow-up questions or provide answers (if you know the answers) as appropriate.

8. Thank the student, again, for participating. Let them know they are always welcome to email any additional comments or suggestions for the course to imanly62@math.ksu.

9. Stop the recorder.

10. Fill out the receipt. Remember to put the wrap around cover behind the receipt. You need to get the students address and social security number. Since we are paying the student, we are legally obligated to get their social security number. Be sure they sign the receipt. Once the receipt is signed, given them $15 and thank them again. Place one copy of the receipt in the envelope with the money and leave the other receipt in the receipt book.

11. Listen to the interview on the recorder and write up your notes. Turn the consent form, your notes, and whatever the students wrote on their pad in to Ian Manly. Transfer the recording to the computer system and erase the IC Recorder.

Notes on the Interview:
Student: ________________________
Date/Time: ________________________
Interviewer: ________________________
Definitions Provided:
Problems Provided:

Problems Attempted:
Comments on the Interview:

Interview Questions

Table A.1: *Cluster - Student Key*

Student	Cluster
1	2
2	2
3	2
4	2
5	2
6	2
7	3
8	2
9	2
10	3
11	2
12	2
13	1
14	2
15	2
16	1

1. What do you think math is all about?

   "Numbers" - 3,5,7,8,11

   "Physics/understanding universe" - 4,11

   "Solving problems" - 1,3,5,7,10,12,14,16

   "Understanding world/how things work" - 2,4,6,8,10,14

   "Applications" - 2,9,15

   "Improving Critical Thinking" - 13

2. Why do we spend so much time learning about math?

   "Used in other areas of study" - 1,2,3,4,6,7,8,9,10,11,13,14,16

   "To solve problems" - 12

   "Useful for society" - 1,2,12

   "Useful in everyday life" - 3,4,5,7,9,10,15

"Develop thinking skills" - 6,13

3. What were your feelings towards math at the beginning of the course? "I took calculus before" - 4,5,6,9,11,15,16

   "Nervous/worried" - 1,8,12

   "I like math" - 2,13,15

   "Math isn't my subject" - 3,7,10,12,14

   "I feel unprepared for it" - 1,2,5

4. Have your feelings changed in the past few months? "Feeling better" - 12

   "Enjoying it" - 1,2,3,5,6,15

   "Difficult" - 2,7,8,10,14,15

   "Feelings haven't changed" - 4,9

5. Are you doing well in the course?

   "Yes" - 3,4,5,6,8,9,10,13,15,16

   "I'm doing ok" - 1,2

   "No" - 14

   "I have an A so far" - 1,3,5,7,8,9,10,13,16

   "I have a B so far" - 4,12

6. Is Calculus 1 what you expected it to be?

   "Yes" - 4,5,6,7,8,9,10,12,13,14,16

   "I didn't know what to expect" - 12

   "I thought it would be hard" - 3,4

   "Different from high school" - 9,14,15,16

7. How much time per week do you spend outside of class for Calculus 1?

   "2-3 hours" - 2,5,8,10,11,14,15

   "Depends on the week" - 12

   "8-10 hours" - 1

   "A few hours" - 3,7

   "5-6 hours" - 4,13

   "3-4 hours" - 6,9,16

8. How do you study for the midterms?

   "Go over the practice exams" - All students

9. If you get stuck on a problem or have trouble understanding a difficult concept, what do you do?

   "Ask friends" - 2,4,5,6,8,10,11,12,16

   "Go to tutoring" - 4,5,11,15

   "Internet" - 2,5,6,12

   "Recitation Instructor" - 4,6,9,15,16

   "Read the book" - 6,14

   "Check notes" - 7

   "Rethink problem" - 13

10. What are your future career goals?

    "Engineer" - 2,4,5,6,8,9,10,11,15,16

    "Pilot" - 12

    "Computer Science" - 1

    "Health Science" - 3,7

    "Teacher" - 13,14

11. What part of the course did you find to be the most helpful?

    “Written homework” - 4,5,11

    “Lecture (all mentioned the same particular instructor)” - 1,9,10,12

    “Recitation” - 1,4,6,7,8,13,15,16

    “Tutoring (not official part of course) - 3

12. What part of the course did you find to be the least helpful?

    “Online homework” - 2,7,8,11,12,15,16

    “Recitation” - 5,10,14

    “Nothing” - 1,14

    “Short exam time” - 3

    “Lecture” - 4,5,6,13

    “Written homework” - 9

13. After answering these quesitons, students were ask to give a conceptual definition of a limit. There answer was...

    “Good/mostly good” - 1,4,10,14,15,16

    “Not good” - 3,5,6,7,9,11,12,13

    “Didn’t try/said they didn’t know” - 2,8

Spring 2017

Rather than in-person interviews, students completed an online questionnaire for the Spring of 2017. Among the 43 students who submitted usable responses, there were six from Cluster 1, three from Cluster 2, 28 from Cluster 3, and six from Cluster 4. The results are given below:

1. Why did you choose to attend college?

Table A.2: *Responses to "Why did you choose to attend college?"*

Response	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Total
Greater Earning Potential	3	2	23	5	33
Knowledge	4	2	22	3	31
Family Pressure	1	0	6	2	9
Other	1	2	2	1	6

2. What are your future career goals?

   Cluster 1 - Artificial Intelligence, NASA, Business Owner, Architectural Engineer, Not sure, Industrial Engineer, Life Science

   Cluster 2 - Building Designer, Restaurateur, Author, Get a job, Not sure

   Cluster 3 - Agricultural Economics, Nanotechnology, Food Science, Grain Science, Pharmacy, Power Systems Engineer, Engineering (3), Surgeon, Mechanical Engineer, Software Developer (2), Software Engineer, Chemical Engineer, Make money, Pediatrician, Airforce, Orthopedic Surgeon, Work for Center of Disease Control, Environmental Engineer (2), Electrical Engineer, Agriculture, Food Safety, Not sure, Medical School

   Cluster 4 - Build machines, Programmer for Air Force, Software Engineer, Work at famous company, Professional Athlete

3. Have you taken Calculus 1 prior to this semester at K-State?

Table A.3: *Responses to "Have you taken Calculus 1 prior to this semester at K-State?"*

Response	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Total
Yes	2	1	4	3	10
No	4	2	24	3	33

4. Did you take a calculus class in high school?

Table A.4: *Responses to "Did you take a calculus class in high school?"*

Response	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Total
Yes	1	0	7	1	9
No	5	3	21	5	34

5. What is your major?

   Cluster 1 - Mechanical/Nuclear Engineering, Architectural Engineering, Computer Information Systems, Industrial Engineering, Undecided, Electrical Engineering

   Cluster 2 - Architectural Engineering, Milling Science, Mechanical/Nuclear Engineering

   Cluster 3 - Agribusiness, Mechanical Engineering (2), Food Science (2), Grain Science, Biology, Electrical Engineering (3), Civil Engineering, Biochemistry (3), Computer Science (5), Chemical Engineering, Industrial Engineering, Psychology, Mechanical/Nuclear Engineering, Microbiology, Environmental Engineering, Geology, Milling Science

   Cluster 4 - Computer Engineering (2), Computer Science (2), Math, Microbiology

6. What year of college are you in?

Table A.5: *Responses to "What year of college are you in?"*

Response	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Total
Freshman	5	2	16	3	26
Sophomore	1	1	6	2	10
Junior	0	0	2	1	3
Senior	0	0	2	0	2
"Super Senior" or Graduate Student	0	0	2	0	2

7. When confronted with a difficult problem in school, what do you do?

Table A.6: *Responses to "When confronted with a difficult problem in school, what do you do?"*

Response	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Total
Work/Try Harder	3	2	8	2	15
Textbook/Notes	0	0	4	0	4
Get Help From Campus Resource	1	0	0	0	1
Use Internet	1	1	7	1	10
Ask Friends	0	1	2	0	3
Ask Instructor	0	0	2	0	2
Tutor	0	2	3	0	5
Other/"Ask for help"	2	0	12	4	18

8. Do you consider yourself to be a good test taker?

Table A.7: *Responses to "Do you consider yourself to be a good test taker?"*

Response	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Total
Yes	2	2	16	1	21
No	4	1	12	5	22

9. What do you hope to get out of Calculus 1?

   Cluster 1 - Find out how world works, Calc 2 prep (2), An A, Pass Calc 1 (2), Course credit, Calculus knowledge

   Cluster 2 - Pass Calc 1, Calculus knowledge, (The other student wrote a page long rant about how bad Calculus 1 is)

   Cluster 3 - Calc 2 prep (5), A B or better (2), Pass Calc 1 (4), An A (7), Learn applications of calculus (3), Required for major (2), Improve math skills (2), Knowledge of calculus (5), Problem solving skills, Course credit

   Cluster 4 - Learn applications of calculus, Knowledge of calculus, An A, Calc 2 prep (2), Pass Calc 1

Fall 2017

Consent Form - Pre-Calculus "Bailout" Course - Why Does it Work?

September 12, 2017

You are being asked to take part in a research study on K-State's Pre-Calculus course. We are asking you to take part because you are currently enrolled in, or have previously taken Precalc at K-State. Please read this form carefully and ask any questions you may have before agreeing to take part in the study.

What the study is about: This study is about understanding Precalc in more detail.

What we will ask you to do: If you are willing, we will conduct an interview with you. We will ask you questions on your opinions of the course and mathematics in general. The entire interview should take no longer than 30 minutes. With your permission, we would also like to tape-record the interview.

Risks and benefits: We do not anticipate any risks to you participating in this study other than those encountered in day-to-day life. There are two benefits - the compensation (described below), and the results of the study, when it is complete.

Compensation: Upon completion of the interview, you will receive $15 cash. If you decide to terminate the interview before its completion, you will not receive the compensation. Also, you will be informed of the results of this research project via email when it is completed.

Your answers will be confidential. The records of this study will be kept private. In any sort of report we make public we will not include any information that will make it possible to identify you. Research records will be kept in a locked file; only the researchers will have access to the records. If we tape-record the interview, we will destroy the tape after it has been transcribed.

Taking part is voluntary. Taking part in this study is completely voluntary. You may skip any questions that you do not want to answer.

If you decide not to take part or to skip some of the questions, it will not affect your current or future relationship with Kansas State University, or your grade in Precalc. If you decide to take part, you are free to withdraw at any time.

If you have questions: The researchers conducting this study are Ian Manly and Prof. Andrew Bennett. Please ask any questions you have now. If you have a question later, you may contact Ian Manly at imanly62@math.ksu.edu or at (315) 935-5099. You can reach Prof. Bennett at bennett@math.ksu.edu or (785) 532-0562. If you have any questions or concerns regarding your rights as a subject in this study, you may contact the Institutional Review Board (IRB) at (785) 532-3224 or access their website at https://www.k-state.edu/comply/irb/.

You will be given a copy of this form to keep for your records.

Statement of Consent: I have read the above information, and have received answers to any questions I asked. I consent to take part in the study.

Your Signature: ____________________ Date: __________

Your Name (printed): ______________________________

In addition to agreeing to participate, I also consent to having the interview tape-recorded.

Your Signature: ____________________ Date: __________

Signature of person obtaining consent: ____________________

Date: __________

Printed name of person obtaining consent: ____________________

Date: __________

This consent form will be kept by the researcher for at least five years beyond the end of the study.

Interview Procedures

September 24, 2017

1. Prepare for the interview at least 5 minutes before the scheduled time. Unlock the conference room (Reta has the key) and leave the door open. Set out the IC Recorder, two copies of the Informed Consent form and a pad of paper for students to write or draw on as needed when they answer the questions.

2. When the student arrives, introduce yourself and welcome the student by name. Close the door to the conference room. Ask for permission to record the interview. If permission is granted, start the recorder.

3. Explain the purpose of the interview:

   We are interviewing students who have taken or are taking Precalculus to better understand the experiences of these students. This is prompted by a desire to understand retention rates both in STEM fields and in the university as a whole. This interview should take approximately 20 minutes. Your participation is completely voluntary and your grade will not be affected by your answers in this interview. You will receive $15 for your time for participating in this interview and you may also benefit by improvements in math course offerings. In the event we include any of your comments in a discussion or publication about our findings, your privacy will be maintained by the use of a pseudonym. We have two copies of an Informed Consent Form for you to sign, one for our records and one for you to keep.

4. Have them read and sign the form. If they decline to sign the form, thank them for their time and terminate the interview. Otherwise sign and date the form as witness and then proceed to the questions below.

5. **Questions**

   (a) *What do you think math is all about?*

(b) *What was your opinion of math before taking Calculus 1? How about while you were enrolled in Precalc?*

(c) *What was your major before enrolling in Precalc? How about after?*

(d) *Which year of college were you in while taking Precalc?*

(e) *Had you taken any calculus before the semester you enrolled in Precalc (High school, college level,etc)?*

(f) *Desrcrbe your experience with Calculus 1 before switching into Precalc.*

(g) *What made you decide to switch classes?*

(h) *How would you describe your experience in Precalc?* Ask the following prompts to get a more specific answer to this question: *How do you think the different format (from Calc 1) affected the course? Do you feel like the class helped you improve your math skills? Did you miss having a recitation like in Calc 1, or do you feel as if you got enough personal attention from the lecturer?*

(i) *What grade did you recieve in the course?*

(j) *Have you taken any math courses since Precalc?*

(k) *Why did you leave the university?* Only ask if the student left Kansas State.

(l) *On a scale from 1 to 5, how significantly did Precalc impact your college experience?*

This concludes our interview. (Thank the student for participating) At this point, read the debriefing statement to the student. If the student has no questions on the debriefing statement, say *Before you leave, Id like to know if there were any questions I should have asked you, but I missed.* Ask follow-up questions or provide answers (if you know the answers) as appropriate.

6. Thank the student, again, for participating. Let them know they are always welcome to email any additional comments or suggestions for the course to imanly62@math.ksu, as well as any comments or concerns regrading this interview to the Chair of the Internal Review Board, Rick Scheidt at rscheidt@ksu.edu or 785 - 532 - 1483.

7. Stop the recorder.

8. Provide student with a copy of the debriefing statement. Read aloud the contents of the statement, and verify that the student understands and accepts the contents.

9. Fill out the receipt. Remember to put the wrap around cover behind the receipt. You need to get the students address and social security number or Wildcat ID. Be sure they sign the receipt. Once the receipt is signed, given them $15 and thank them again. Place one copy of the receipt in the envelope with the money and leave the other receipt in the receipt book.

10. Listen to the interview on the recorder and write up your notes. Turn the consent form, your notes, and whatever the students wrote on their pad in to Ian Manly. Transfer the recording to the computer system and erase the IC Recorder.

Notes on the Interview:
Student: ______________________________
Date/Time: ______________________________
Interviewer: ______________________________
Comments on the Interview:

Debriefing Procedures

September 24, 2017

At the conclusion of the interview, the following debriefing statement will be read to the participant:
Thank you for participating in this interview. The purpose of this study is to better understand the Precalculus "bailout" course and its effectiveness. In particular, the researchers would like to know what aspects of a "bailout" pre-calculus class are most important in creating conditions for student success and retention. We would like to find out how can we better support students who struggle in their initial math class so they are ultimately successful. The primary risk of your participation in this study are having your identity linked to your grades in Precalc. This risk is minimized by the use of a numbering system, as to not link your identity to your scores. Any information from this interivew used in any reports or papers will not be attributed to your name or any other identifying factors. Other than this, there are no anticipated risks other than those encountered in daily life. Since you have participated in this study, the benefits you will recieve are the monetary compensation of $15, as well as improvements to mathematics instruction. Upon completion of the study, you will be emailed a report with the study and any findings.

Interview Questions

In total, there were 18 Precalculus students who were interviewed. Students 1-6 had taken the course in a semester prior to the Fall of 2017, whereas the remaining 12 were students from the Fall of 2017. Below are the questions the students were asked, and what sorts of phrases they said or responses they gave.

1. What do you think math is all about?

   "Weird question" - 1

   "Used in calculations and problems" - 1,2,3,4,5,8,12,13,15,17,18

   "Applied in other fields" - 1,4,10,12

   "There's "more to it"" - 2

   "Don't have a good answer" - 2

   "Understand the world better" - 4,8,13,14

   "Used in day to day life" - 4,7

   "Making life more difficult for me" - 6

   "Language" - 10

   "Numbers" - 6,8,11,18

   "To be memorized for class" - 9

2. What was your opinion of math before taking Calculus 1?

   "Positive opinion of it in high school" - 1,6,12,15,18

   "Got worse in college" - 1,3,7,10

   "Hated it" - 2,17

   "Means to an end/require course" - 2,4

   "Difficult" - 2,5,9,14,17

   "It's okay" - 3,4,11

   "Easy" - 8,10,13

3. What was your opinion of math while enrolled in Precalculus?

   "Got better since the class was easier" - 1,2,3,6,7,8,9,11,17

   "Got more interesting" - 2

   "Enjoyed practicing problems" - 3

   "Liked the detailed explanations" - 14,18

   "My opinion was the same as before" - 4,12,15

   "I gained more confidence" - 5,13

   "A lot of memorization" - 8

4. What was your major before enrolling in Precalculus? How about after?

   Civil Eng. → Construction Management - 1

   General Eng. → Math/2nd. Education - 2

   Biochem. → Biochem. (w/ nursing concentration) - 3

   Biochem. → Biochem. - 4

   Kinesiology → Zoology - 5

   Bio Eng. → Accounting - 6

   C.S. → C.S. - 7,13

   Mech. Eng. → Mech. Eng. - 8,11

   C.S. → I.S. - 9,14,15

   Chem. → Life Science - 10

   Physics → Business (Undecided) - 12

   Geology → Geology - 17

   Civil Eng. → Civil Eng. - 18

5. Which year of college were you in while you were taking Precalculus?

   Freshman - 2,4,5,6,7,8,11,13,18

   Sophomore - 1,9,10,14,15

   Junior - 3,12,17

6. Have you taken any calculus prior to the semester you took Precalculus?

   Yes - 5,11,13,18

   No - 1,2,3,4,6,7,8,10,12,14,15,17

7. Describe your experience with Calculus 1 before switching into Precalculus.

   "Didn't like how it was run" - 1,7,13

   "Went too fast" - 1,5,6,7,10,12,14,15,17

   "Liked recitation more" - 1,2

   "Terrifying/intimidating" - 2,17

   "Couldn't ask questions" - 2,5,11

   "Did poorly on the first exam" - 3

   "Difficulty understanding concepts" - 4,8

   "Not prepared" - 5,6,8,10,18

   "Felt intimidated by people who already knew calculus" - 5,17

   "Got harder as semester progressed" - 9

   "Not as straightforward as previous math classes" - 13

8. What made you decide to switch into Precalculus?

   "Did poorly on the first exam" - 1,2,3,6,8,13,15

   "Wanted to replace grade" - 1,3,5,10,14,15,18

   "Overwhelmed with Calc 1" - 2,4,17

   "Advisor recommended" - 4

   "Wanted to stick with major" - 4 "Wanted to get better at calc and math" - 5,7,9,11,13,14,15,17,18

   "Didn't need it for new major" - 10

   "Wanted to focus on other classes" - 12

9. Describe your experience in Precalculus.

   "Liked it" - 1,2,5,7,14,17

   "Had a good teacher" - 1,12,15

   "Good review" - 2,6,8,12,15

   "Wanted something in between precalc and calc 1" - 2

   "Got to learn why certain things are true" - 11,12

   "Wished there was a textbook" - 15

10. What did you think of the format of Precalculus?

   "More personal" - 1,2,5,7,8,10,11,12,15,18

   "Good speed" - 1,2,8

   "Got to work on other problems" - 3,5

   "Liked meeting fewer times per week" - 3,7

   "Reminded me of high school" - 4,11,13,17

   "Got to work with others" - 6

   "Missed old format" - 8

   "Got to ask questions" - 14

11. Do you think taking Precalculus improved your math skills?

   "Yes" - 2,3,4,9,10,11,12,17,18

   "Helped with understanding calculus" - 2,3,4,6,7,10

12. Do you miss having recitation like in Calculus 1?

   "No, I got enough attention in class" - 1,2,4,5,6,9,13,18

   (interviewee called Bailout Precalculus class meetings "recitations") - 2,3

   "No, recitation wasn't helpful" - 5,7,10,11,14,17

   "Yes" - 8,12

13. What grade did you receive in Precalculus? (only asked of students from previous semesters)

    "A" - 2

    "B" - 1,3,4,5,6

14. Have you taken any math courses since Precalculus? (only asked of students from previous semesters)

    "Business Calculus" - 1

    "Calculus II" - 2

    "Calculus I" - 2,3,4

    "No" - 5

15. On a scale from 1-5, how significantly did Precalculus impact your college experience?

    "2" - 15

    "2.5" - 6

    "3" - 9,17

    "3.5" - 2

    "4" - 3,4,5,10,14,18

    "5" - 1,11,12,13

Appendix B

Additional Data Mining Figures

This Appendix contains additional singular value plots, dendrograms, and PAM plots from configurations the researcher considered, but elected to not use in the Fall of 2016 and the Spring of 2017. See attached file for a sample of the R code used to data mine and produce these diagrams.

Fall 2016

The researcher determined that using a single entry for total attendance yielded the most sensible clusters for analysis. However, using the daily attendance as individual entries and looking at weekly attendance scores also produced somewhat viable clusters.

For the individual attendance version, the researcher used 3 dimensions of data and 3 clusters, as shown in figures B:1-6. For the weekly attendance version, the researcher decided to use 4 dimensions of data and 4 clusters, as shown by the following figures:

Spring 2017

The researcher determined that using each day of attendance as an individual entry yielded the most sensible clusters for analysis. However, using one entry for the total attendance and looking at weekly attendance scores also produced somewhat viable clusters.

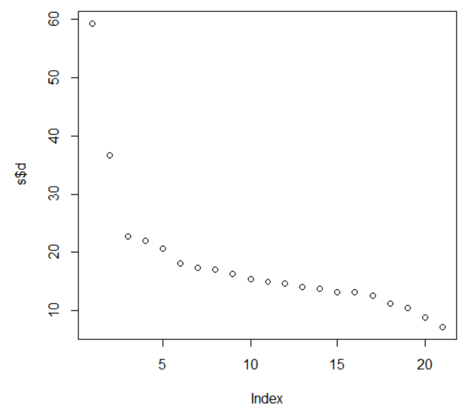

Figure B.1: *Singular Values for the Fall of 2016 with individual attendance used.*

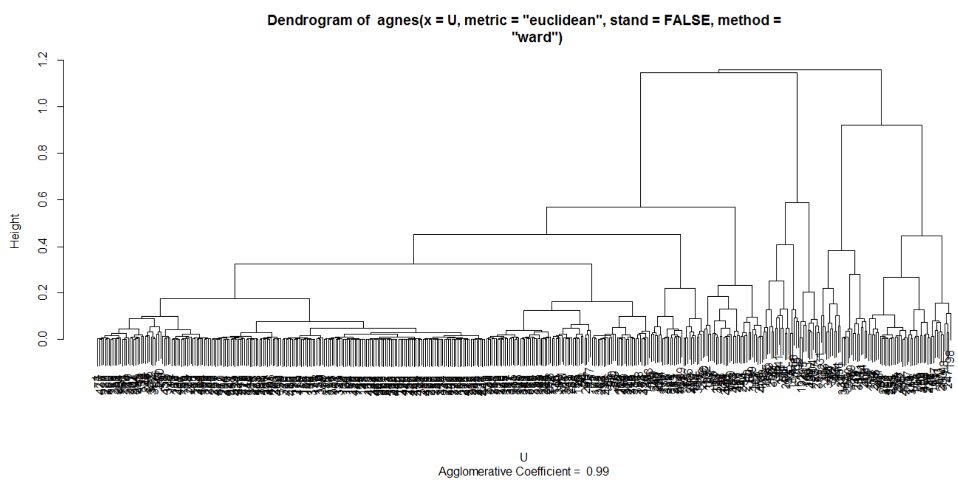

Figure B.2: *Singular Values for the Fall of 2016 with individual attendance used.*

For the total attendance version, the researcher decided to use 4 dimensions of data and 4 clusters, as shown by the following figures B.7-12.

For the weekly attendance version, the researcher decided to use 4 dimensions of data and 4 clusters, as shown by the following figures:

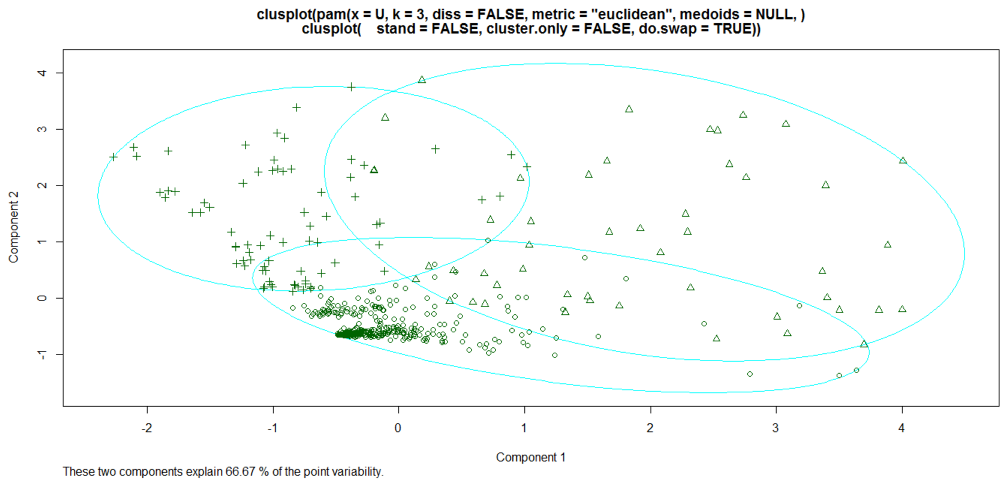

Figure B.3: *Singular Values for the Fall of 2016 with individual attendance used.*

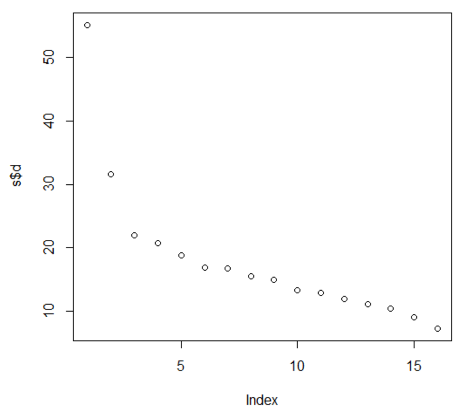

Figure B.4: *Singular Values for the Fall of 2016 with weekly attendance used.*

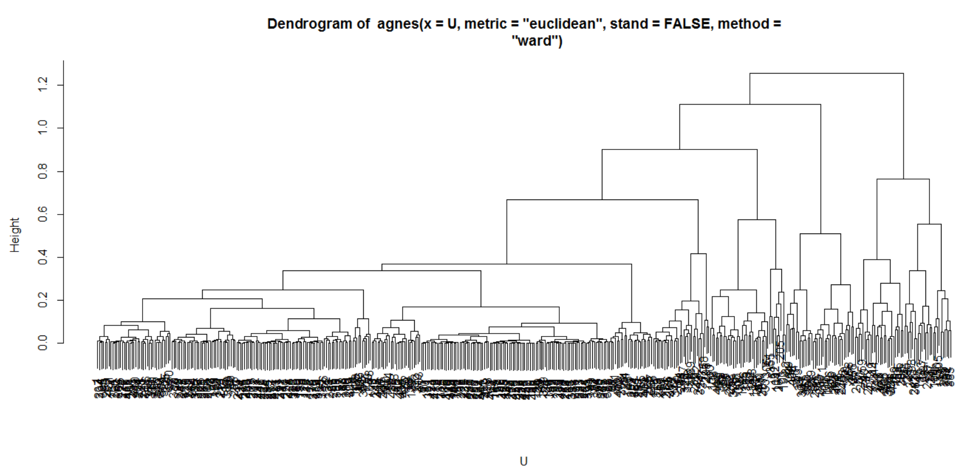

Figure B.5: *Singular Values for the Fall of 2016 with weekly attendance used.*

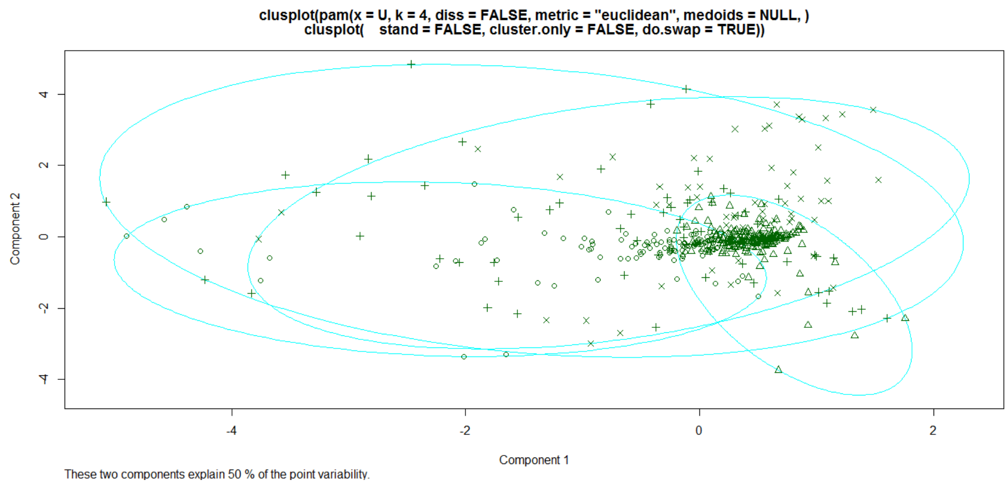

Figure B.6: *Singular Values for the Fall of 2016 with weekly attendance used.*

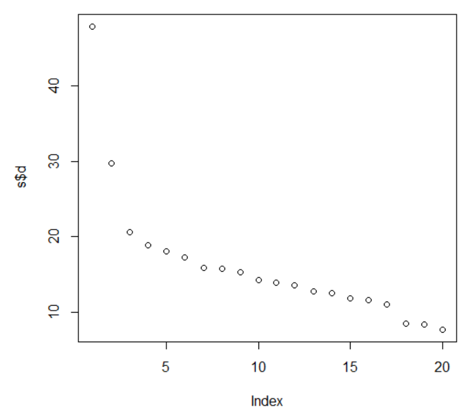

Figure B.7: *Singular Values for the Spring of 2017 with total attendance used.*

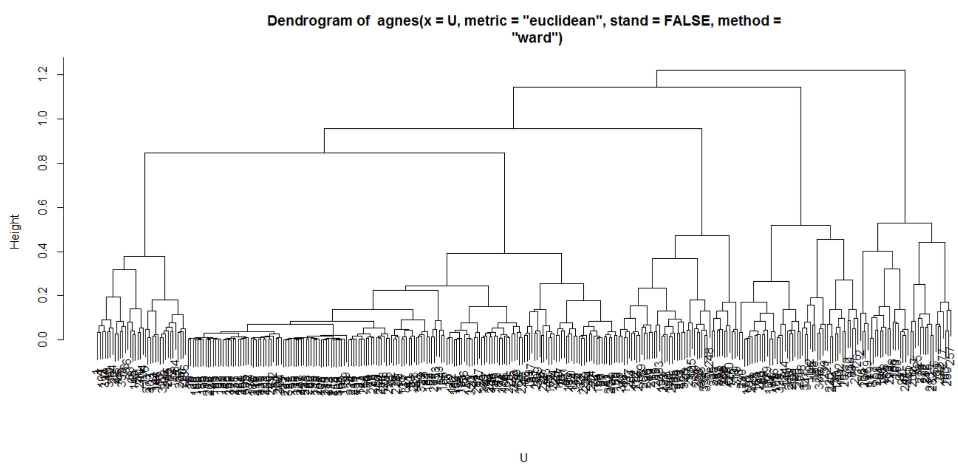

Figure B.8: *Dendrogram for the Spring of 2017 with total attendance used.*

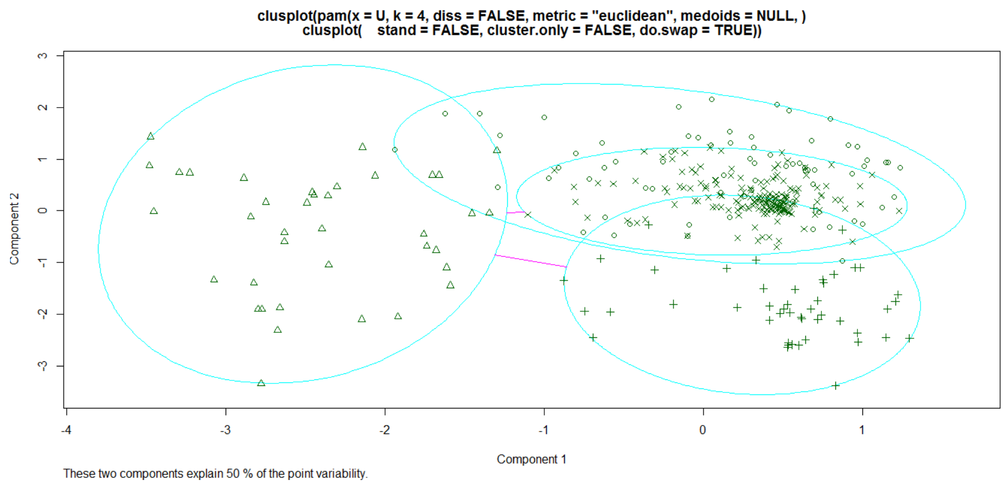

Figure B.9: *PAM plot for the Spring of 2017 with total attendance used.*

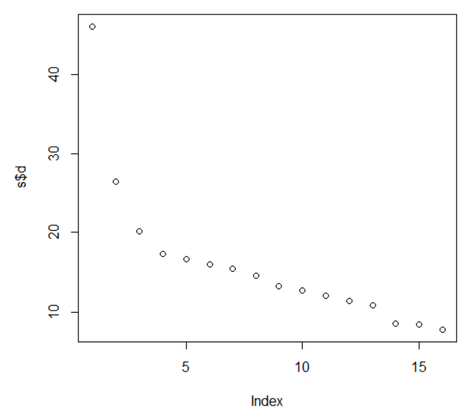

Figure B.10: *Singular Values for the Spring of 2017 with weekly attendance used.*

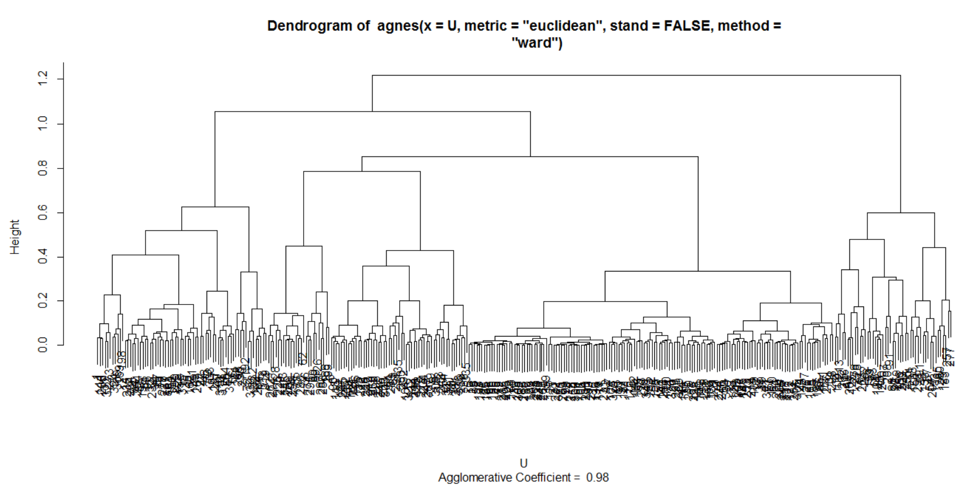

Figure B.11: *Dendrogram for the Spring of 2017 with weekly attendance used.*

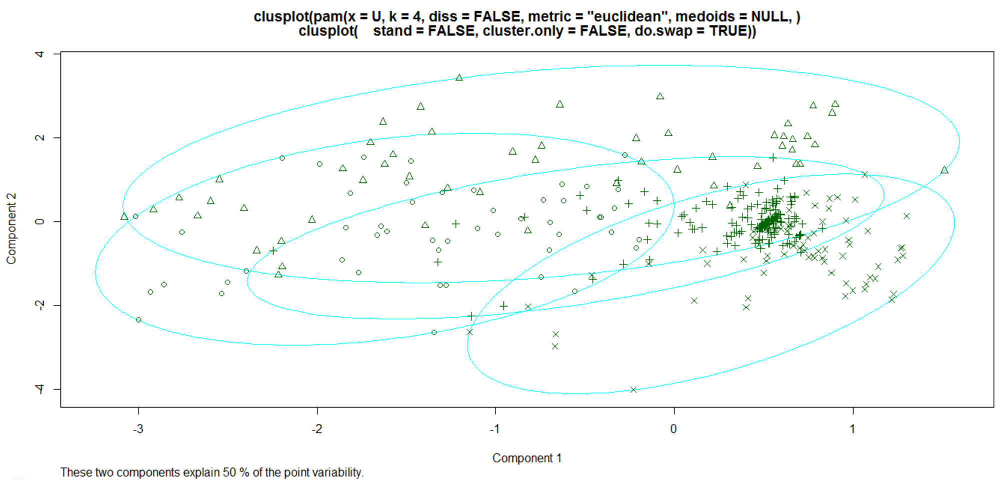

Figure B.12: *PAM plot for the Spring of 2017 with weekly attendance used.*