INDIANA UNIVERSITY–PURDUE UNIVERSITY INDIANAPOLIS

A REPORT SUBMITTED IN PARTIAL FULFILLMENT FOR THE DEGREE OF MASTER'S

In the
Department of Physics
IUPUI

Data Mining in Physics Education Research

Patrick Kelley

supervised by
Dr. Gavrin

April 6, 2022

Contents

Chapter 1

Introduction

1.1 Preface

The Indiana University – Purdue University Indianapolis (IUPUI) introductory physics course, labeled PHYS 152, implemented a social media application developed at IUPUI called CourseNetworking (CN). CN has a similar layout to Facebook and likewise is a free online social networking platform designed to compliment courses and provide a means for students to socialize outside the classroom. CN has the capability to store information on their social activities, having recorded four semesters of students' online discussions in PHYS 152. With this large set of data and few researchers to look it over, a means of a computer automated analysis technique was strongly desired. It turns out that making sense of big data is a rapidly rising focus for many researchers, who obtain much larger swaths of data. A popular track for analyzing textual data, called text mining, is either through the open-source programming languages Python or R. Because of its statistical nature, we decided to utilize R. The way we approached it is by finding and installing packages for different data analysis. Our first approach, after visualizing aspects of the data using packages that generate word clouds and network plots, was in the form of sentiment analysis. We heavily made use of Syuzhet, an R sentiment analysis package. The way Syuzhet was designed was to match words from the online discussions and use one of four available lexicons comprised of a selected list of words with ascribed sentiments, whether positive, negative, or none, and additionally eight emotions. The weight of sentiments and emotions to each selected word from the independent lexicons were crowdsourced. With this at our disposal, we broke each student post and reply into words, running those through Syuzhet to form a matrix of quantitative sentiment and emotion values. We could then look at the distribution of frequency of postings and gender versus the resulting sentiment analysis.

We continued with our investigation of this data after publishing a conference preceeding paper at the Summer 2017 Physics Education Research Conference (PERC) in Sacramento, CA. Our current focus targets the temporal aspects of students' posting. We looked at the perspective, often held by faculty, that late-night students perform differently academically than consistent day-time students.

The ultimate goal of the study is to gain deeper understanding of students in order to promote a better learning experience. National, state, and university institutes as well as individual researchers express desire for studies devoted to this same goal, prompting a need for change in education practices for STEM fields [1]. A web based survey reports [2] that many instructors respond positively for research-based educational innovations, however express lack of time to investigate and incorporate these tools and instruction practices in their own course. A tool such as CN that is easy to set up and monitor would greatly benefit instructors in understanding their students to a much greater degree: many who may have responded to the survey that student attitudes and appreciation for physics ranked low in their teaching importance. CN could provide a stepping-stone for instructors who have trouble changing to or adopting correct research-based innovations, the way researchers have experimentally and rigorously proven improves learning. The motivation for CN corresponds with many of the general characteristics of learning improvements, most specifically to student-student interactions, encouraging peer-to-peer active learning that can lead to better conceptual learning.

It may shift the attitudes of instructors who declared high willingness to try research-based teaching practices in the survey yet report high levels of satisfaction on reaching these learning improvements although they continue to use traditional instruction practices. Research demonstrates fewer improvements in respect to student attitudes, deeper concept understanding and longer knowledge retention, such as in the case of active learning versus traditional instruction [3].

In brief, CN provides an insight to how students perceive the course and fosters easier communication not only among students but also between students and the instructor, whereby some students may find it simpler to post something on CN rather than draft a direct email to the instructor or find time to meet with the instructor in person. This modernized openness can allow instructors to evaluate their own instructional practices and adjust their own views on the learning experience of their students, in order to better facilitate a more effective educational instruction.

We will begin by laying out the physics course framework, describe the social media platform implemented and detail the motivation with literature review. The purpose of this chapter is to paint a picture of how the course is structured, what this web tool is and the data it collects from students. The next chapter thereafter will explain the approach using R with a breakdown of the code. The results are then presented in the following chapter. We report lastly with a discussion of the results and the methods and possible future directions of our research.

1.2 Course Setting

IUPUI is largely a commuter school. Student-student interactions are thereby limited to class times or private social media platforms that students form independently. To give students easier transparent communication that can be monitored, classes can utilize a free social networking platform for academic use called CourseNetworking (CN). This is what has been done for four and continuing semesters of PHYS 152, an introductory physics course at IUPUI.

The introductory physics course is four credit hours; students attend two 2-hour lectures, one recitation hour and a laboratory hour in the week. Students complete homework on an online platform that counted for 15% of students' course grade. Before fall 2015, it was Macmillan Learning FlipItPhysics. From fall 2015 onward, students use Webassign. In addition, the eText Physics for Scientists and Engineers, 6th edition, by Tipler and Mosca comes complementary with homework assignment access on Webassign, making the mandatory traditional hardcopy textbook optional. Exams were administered thrice throughout the semester before Fall 2015, but was increased to 5 exams distributed equally throughout the course. Overall, the exams constituted 30% of the students' course grade with a 20% comprehensive final exam administered at the end of the course.

Lectures are held twice a week for 50 minutes each; with Turning Technologies tracking class participation via RF clickers, a personal response device purchased before the beginning of the course. The clicker questions, given in lectures and recitations, makes 10% of students' course scores. However since Fall 2018, these Turning Technologies RF clickers have been replaced with TopHat, an educational software company that offers an application accessible via smartphones and laptops with in-class engagement and real-time feedback functionality. All lectures implemented Just-in-Time Teaching (JiTT) [1] and Peer Instruction (PI) [2].

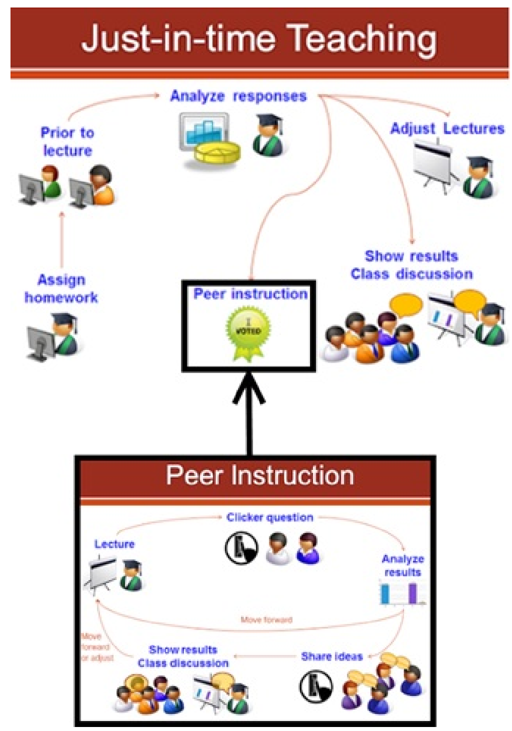


Figure 1.1: An illustration of JiTT

JiTT is mixing active learning with web technology, namely the predominantly wide spread use of online homework and other online classroom tools. Figure 1.1 outlines a simple process in which the instructor gives online assignments, called warmup exercises, that ask students to prepare before class by reading and answering a few questions. This exercise constitute 10% of students' overall grade in PHYS 152. This task is performed and tracked on Webasign. The instructor assesses the responses and results from the students and can do three things in lecture. First they modify their day's lecture to accommodate subjects that students appear to be deficient in from the pre-class results. Secondly, the instructor can reveal the source of difficulty to the class, hold discussions and active interactions in the classroom. Lastly, the instructor can use peer instruction, a teaching practice popularized by Harvard professor Eric Mazur. The instructor would pose a question in class from a PowerPoint presentation, prompted by common pre-known conceptual difficulties or by students' pre-class assignment results. Students would be given time to individually assess the question and provide an answer in the form of multiple choice options submitted with clickers. They would then have another attempt to answer the question via clicker, after students had some time to discuss amongst their fellow classmates and share ideas on the question. Based on the results, the instructor can progress with the lecture or spend extra time on the concept the students find difficulty in.

The basic concepts that are covered in this calculus-based mechanics course are as follow-

Table 1.1: Course outline for six semesters of PHYS 152

	Text Book	**Clicker**	**PI**	**JiTT**	**Number of Exams**	**HW System**	**Year of CN in PHYS 152**
Fall 2014	SmartPhysics[a]	TT[b]	✓	✓	3	SmartPhysics	1
Fall 2015	*PSE* [c]	TT	✓	✓	5	WebAssign	2
Fall 2016	*PSE*	TT	✓	✓	5	WebAssign	3
Spring 2017	*PSE*	TT	✓	✓	3	WebAssign	4
Fall 2017	*UP* [d]	TopHat	✓	✓	5	WA & MP [e]	5
Fall 2018	*UP*	TopHat	✓	✓	5	WA & MP	6

[a] now called FlipItPhysics
[b] Turning Technology (TT)
[c] Tipler & Mosca *Physics for Scientists & Engineers (PSE)*
[d] Young & Freedman *University Physics (UP)*
[e] WebAssign (WA) & Mastering Physics (MP)

ing: uniform and accelerated motion; Newton's laws; circular motion; energy, momentum, and conservation principles; dynamics of rotation; statics, gravitation and planetary motion; hydrostatics and hydrodynamics; simple harmonic motion; wave motion and sound.

The course consists mostly of engineering majors with approximately 150 students per semester, who have taken courses in mathematics with prerequisite understanding of algebra, trigonometry and calculus. Furthermore, all students come in with their own views and understanding of physics, and it's important to note that students aren't 'blank slates' as often assumed. Those views, influenced by social environment and prior experiences, include attitudes, beliefs, and assumptions about what they will learn and gain in terms of skills or knowledge and how they expect to perform in the course. These views have a strong influence on how the student will do in the course, and ultimately if students learn lasting or superficial knowledge and skills. Instructors tend to hold a list of expectations of the students separate and self-evident from the syllabus, namely:

> "to make connections, understand the limitations and conditions on the applicability of equations, build their physical intuition, bring their personal experience to bear on their problem solving, and see connections between classroom physics and the real world" [3].

But evidently not all students follow the instructors' "hidden curriculum" [3]. When students' are frustrated, confused or simply uninterested in the instructor's teaching, they resolve to more 'efficient' means to complete the course with the only thing that they view matters, the course score. They will learn only algorithmic approaches to problems and brute force memorization of poorly understood equations and formulas. These emotions that prompt student to behave thusly is part of what psychology identifies as affect, the experience of feelings and emotion. Clearly, affect has a major influence on student learning.

1.3 Affect

The role of affect is a hard facet to address in the educational setting; the 'cognitive' aspects, a bit easier to facilitate, has instead prevailed in course development. Whether students would get frustrated or excited, instructors have nevertheless been tasked with covering large amounts of material in many cases, in short amount of time. They have resorted to giving large amounts of information in single lectures, usually test students' attention spans, and assigning online automatically-graded problems that more than often not, does not spark students' interest. All of these practices come at the price of neglecting to take into effect how students are emotionally and motivationally coping in the course. It hasn't been until recently that education has seen a resurgence to study the part that affect plays in learning, especially with the advent of computer technology in the classroom. Even so, no complete validated theory of emotion has up until now surfaced that can account for how emotion influences learning or even yet, which emotions are most important to learning [4].

Affect is taken to be a more extensive form of emotion, expanded to encompass things such as motivation, boredom and other states of mind [5]. I have mentioned its importance but its pertinence can be illustrated by vocal intonation, the modulation of one's voice when speaking. Infants and even dogs can pick up and respond to the mood of the speaker, without ever understanding what was actually said [6]. A similar phenomenon even happens to adults, when listening to politicians or public speakers. The way someone says something may have as much effect as what was said. These effects on emotional states can have a ramification on a persons' cognitive ability. Isen and Means [7] examined subjects that were given six fictitious cars to choose from with variables important to the imaginary purchase. Those subjects that received positive feedback from a prior perceptual-motor task were more likely to make quicker decisions, less likely to review information they previously inspected as compared to the control group, and ignore information deemed unimportant to the car purchase. The conclusion was that positive affect increased efficiency on decision-making.

This seems to extend to social media. The interactions from social media influences affect, which if positive can have some apparently favorable effects. Oh et. al. studied this by recruiting 339 undergraduate student participants from Midwestern university and have them log their social network site activities (Facebook, Twitter, etc). The analysis yielded a positive correlations among the number of SNS friends, supportive interactions, positive affect, perceived social support, sense of community, and life satisfaction [8]. On the other hand, there has also been considerable studies on the distraction of social media in the context of people multitasking between social media and academic tasks. Van der Schuur et al. [9] investigated multitasking between social networking sites and its impact on the three aspects of "youths" (eg adolescents, students, and undergraduate students): cognitive control, academic performance, and socioemotional functioning. They report on the work done by Shih [10] and Becker et al. [11] on undergraduate students with regards to the last aspect, socioemotional functioning. Shih found no correlations between multitasking and emotional well-being, however Becker et al. found problematic emotional relations to multitasking with increased social anxiety and symptoms of depression. Van der Schuur et al. conclude from this work and studies on pre-secondary education youths that social media multitasking relate to less sleep, more sleep issues and diminished emotional functioning.

Although Van der Schuur et al. showed negative correlation for individual emotional functioning, no linkage appears between social functioning and multitasking. This may be due to a sense of caring, which has been shown to be motivating [12] and hints as to why social media has such an appeal. Social media obviously plays a role on affect and affect can influence learning [13, 14, 15]. Ractham and Firpo investigated whether social media networks can enhance learning, by using Facebook [16]. The study centered on 69 students taking a graduate-level Management Information System (MIS) course at Thammasat University in Bangkok, Thailand, as well as 2 lecturers and 1 assistant instructor, producing 2640 posts in this Facebook Social Learning Project. The 2 instructors proceeded to analyze and code all the posts. They observed for students, congenial learning from informal interactions, feedback on thoughts, and support for collaboration independent of space and time. The detected values of Facebook for instructors include constant continuous communication and feedback from students as well as an effective instructional tool in teaching.

Similarly, we introduced a social media platform, called CourseNetworking, as a social experiment for PHYS 152, in the hopes that it would increase positive affect from social support structures and a caring atmosphere, thereby increasing the potential for learning.

1.4 CourseNetworking

CourseNetworking (CN) [17] is a global academic networking site. The founder of CN is Dr Ali Jafari, professor of Computer and Information Technology at IUPUI as well as the director of Cyberlabs located at IUPUI, a research and development site for CN which also holds offices in Kuala Lumpur, Malaysia and Guangzhou, China. The function of CN is to provide a course "networking" platform rather than course "management" system. CN enables students to communicate more efficaciously than traditional learning management systems (LMS). The LMS at IUPUI is Canvas, which has the advantage of having a central hub where course material is retrievable and administered. The only means of communication typically on the LMS is an online discussion or announcement board. CN provides a more natural environment for networking and communication, with many similarities to features implemented by Facebook. An interesting feature of CN is anar seeds which are a form of "micro-rewards". These are gained through use of CN via logging in, posting, reflecting and even polling in addition to clicking a "like" button for each post or reflection similar to Facebook. Another similarity with Facebook is a "wall" format with postings by all the members made visible at the home page of the class or group. The control of the instructor is in the setting up of the class or group page with the point weight decided in advance for each action (login, posting, reflecting, polling and liking). A limit can also be imposed on the number of anar seeds students can gain each day. This hinders students from accumulating all required points for extra credit at the end of the semester. The introductory physics courses awarded 5% extra credit if more than 350 anar seeds were earned.

An example of gaining 350 anar seeds by the end of the course would be 50 logins [1 anar seeds/login], 5 posts [10 anar seeds/post], 15 reflections [5 anar seeds/reflection], 15 polls answered [5 anar seeds/poll] and 50 likes [2 anar seeds/like].

Unlike Facebook, CN is moderated by the instructor which gives the instructor an insight into his/her class that Facebook may prevent when students form closed group with discussions and interactions that the instructor is unaware of.

Figure 1.2 gives an example screenshot of what the platform looks like for a course:

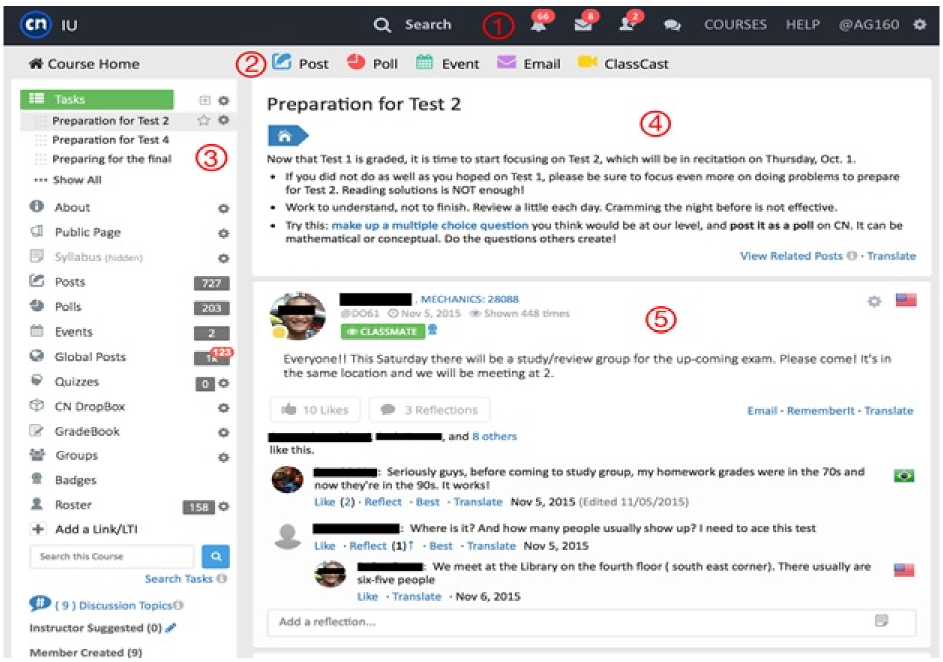


Figure 1.2: Sample view of CN user webpage

1. **CN Navigation Bar** – Logo to left, and going right the Search, Alerts, Mail, Contacts, Chats icons with a link to the user's personal Courses, Help for more information and the @users settings icon.

2. **Course Home Call to Action** – Create a Post, Poll, Event, Email, or ClassCast by clicking on the corresponding icon

3. **Course Home Sidebar** – Tasks created by the instructor, About is an area explaining what the course is all about, Public Page is the site description that is viewed publicly, a syllabus can be uploaded to the Syllabus link, and the following links with numbers will arrange the course based on the categories (Posts, Polls, Events, Global Posts, and Quizzes).

4. **Main Content Headline** – Since 'Preparation for Test 2' was selected under Tasks in **3**, this area displays the outline for the task, written by the designer of the task, namely the instructor.

5. **Content** – Posts and Replies. The first comment is the post and the subsequent comments are the replies to the original post or previous commenter. Another interesting

feature is the name of the student which is blacklined along with their self-declared nationality exhibited by the flag to the right of the name is present to all members of the course.

CN collected all the course posts and reflections along with a corresponding time stamps, number of picture attachments and number of "likes" of each student distinguished by a unique student identification code. CN provided this data for later analysis to the instructor. The data spans the three fall semesters (2014, 2015, and 2016) as well as one spring semester (2017) by a different professor.

Around 3000 posts and reflections were recorded in a semester in the 19 weeks of each course. Fall 2014 opened the course on CN several weeks before the actual start of the semester, with the instructor welcoming the class and inviting students to break the ice by providing some basic introductions. The early self-introduction along with any following posts made by the instructor in the following semesters were later removed to study the student-student interaction and sentiment in the classes. Figure 1.2 outlines some information about course structure and basic statistics about the data collected from CN.

Table 1.2: Course details for the four semesters of PHYS 152

	Instructor	**Instruction Material**	**Number of exams with a final exam**	**Total number of student posts**	**Number of weeks with active posts**	**Range of semester posts per students**	**Number of students who used CN**
Fall 2014	P1 [a]	SmartPhysics	3 exams	3150	24	1-90	151
Fall 2015	P1	WebAssign [b]	3 exams	2971	19	1-122	128
Fall 2016	P1	WebAssign	5 exams	3125	18	1-160	143
Spring 2017	P2 [c]	WebAssign	3 exams	4764	19	1-106	157

[a]Professor 1
[b]WebAssign with E-book version of *Physics for Scientists & Engineers*
[c]Professor 2

1.5 Network Analysis

Although we largely restricted data mining in this report to sentiment analysis, we also performed network analysis briefly at the beginning of our exploration of some of the available tools on R for data mining. We made use of some fundamental network visualization tools in R when first experimenting on our data, before heading in the direction of sentiment analysis. We prepared the data by interpreting reflections as connections to the original poster, or user that created the post thread. We arranged the data in a spreadsheet by creating a column of users, labeled by their unique identification numbers, separated by the student (in the From column) in communication to another student (in the To column).

Then with the **igraph**() function in R, we were able to produce a network plot of the fall 2014 semester, as an example. The initial plot was extremely tangled with their id number as individual nodes. Several functions exist to condense the jumbled mess into something more coherent. The function **walktrap.community**() does simulated short random walks to find packed regions of network that form 'communities'. The following graph illustrates the coagulation of the scattered network from five steps defining a community. The user id with the highest degree, or largest number of edges, became the name of the community node. The function **contract.vertices** then graphically merges the nodes into one and with some tweaking of the graph settings, yields the following igraph plot displayed by Figure 1.4:

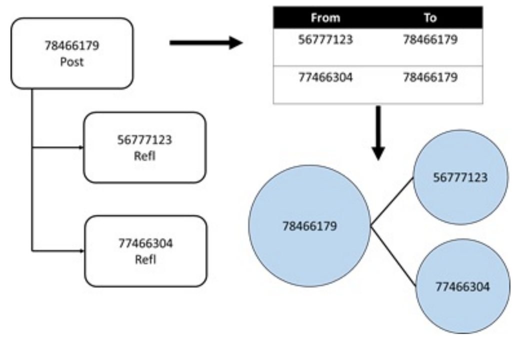


Figure 1.3

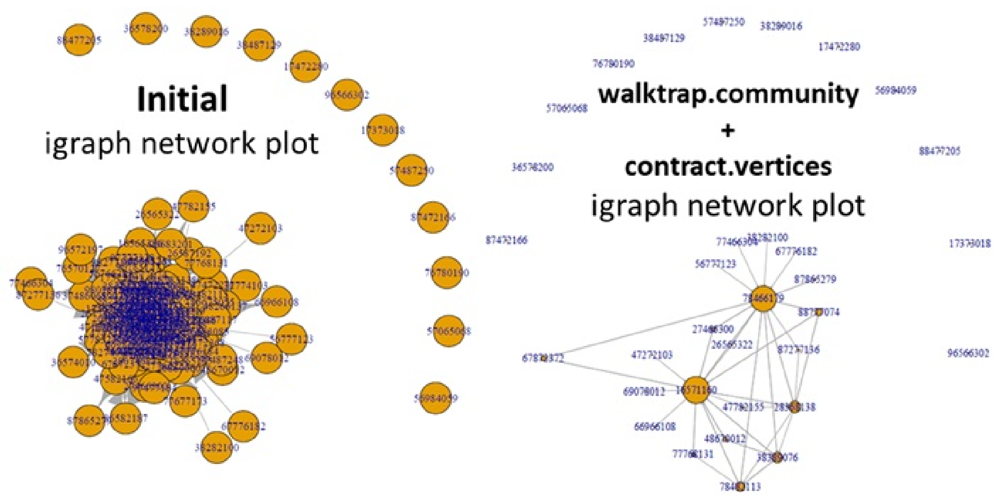


Figure 1.4

Other physics education researchers have carried out in-depth work on the same data. Adrienne Traxler from the Department of Physics at Wright State University in Fairborn, Ohio et. al. works on network analysis on student-student interaction and calculated correlation between the weight of their interaction to their final course grade. She does so by using a bipartite network model, in which students (actors) are linked to each other by the number of common post threads that they create or respond to via post or reflection (events). Figure 1.5 includes a modification of one of her figures from her paper [18] to illustrate how she defined her networks.

This procedure produces an unwieldy amount of interconnections between students; a measure of importance of a node (student) in the network called centrality was used to make sense of the network. Three different types of centrality values were tested (i) a Google search

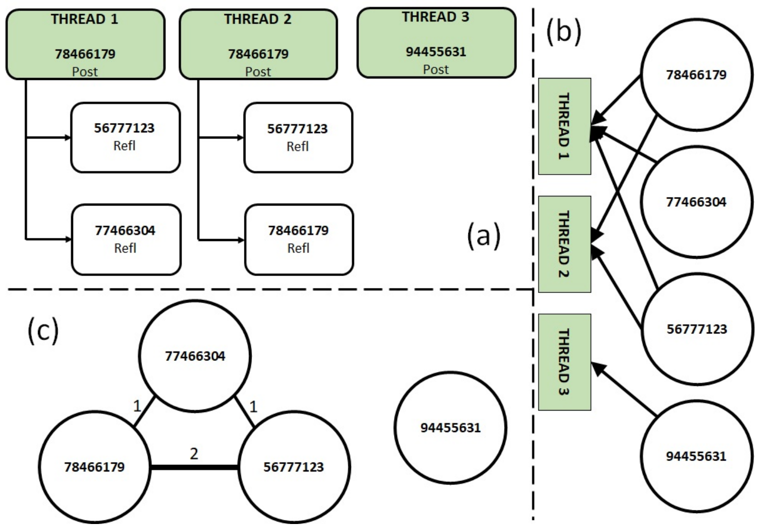


Figure 1.5: Bipartite Network Model implemented by Traxler et al. (a) Original posts are identified as threads (b) The students, encoded by a number, are linked with the threads including the student who created the thread, and (c) the student as nodes and the edges as the students links to the threads and thus each other create the networks

engine based algorithm called PageRank that finds the importance of the node by how many other important or highly weighted linked nodes are connected to it, (ii) Target Entropy that measures the dispersion pattern of students connecting with other students that themselves are mingled with many other students, and (iii) Hide that measures the opposite of Target Entropy, whereby nodes possess a larger value for students with less connections than their average neighboring student CN user. A Pearson correlation is calculated between these three centrality measures and the students' final course grade, with permutation tests of the nodes to account for the violation of network centrality measurements on the basic premise of correlation studies.

From the most active users, there might be a deeper foundation to the seemingly unharmonious clutter of networks. Locally Adaptive Network Sparsification (LANS) algorithm exposes any likely buried structures using a backbone extraction on the overall network. The use of PageRank, Target Entropy and Hide revealed good correlation between centrality and course score but surprisingly LANS provided lower correlation values as compared to these three centrality measures, since LANS supposedly simplifies the messy "noisy" network to discover more solid hidden structures. Traxler et al. drew the conclusion that the "noise" is integral with the data, that deconstructing the network into a simpler backbone representation actually impairs the correlation detected between the centrality and the scores. The finding ultimately revealed that more integral students, those with a more central network position, linked with higher course success.

Networks play a chief means of analyzing data regarding interactions. Although the small sampling I did is crude compared to the work done by Traxler, future implementation of networking could be used alongside sentiment analysis. Possible directions of network analysis in the works will be discussed in more detail in the last section of this report, Future Work.

Chapter 2

Methodology

2.1 Preface

Text mining, also known as Text Data Mining, Intelligent Text Analysis, or Knowledge-Discovery in Text (KDT) is a method used to extract useful information from written words. Text mining, such as sentiment analysis, applies pattern recognition and statistical analysis to natural language, which is often not structured. The power of using these tools is the ability to handle quantities of textual data too large to be manually read and analyzed. Thus, text mining offers a solution that goes beyond numerical data mining. The generalized steps of text mining are to structure the data by parsing, filtering, and collating it, then finding patterns and matching to word databases such as lexicons to obtain qualitative values, and finally evaluating and displaying the output results.

First we will outline how we applied text mining, particularly sentiment analysis (SA), with our selected programming language and with some description on the background to how it functions. We will break down the primary code that performs SA on our data, and how some of the ensuing results were structured to be analyzed. The complete SA code can be found in Appendix A.

2.2 Syuzhet Package and NRC Lexicon: Eight Emotions & Two Sentiments

Three most common or popular avenues for text mining is either with Java, Python or R. In these studies, R was used. R is an open source programming software designed with statistical analysis in mind [19]. It comes with easily downloadable packages. One such package is called Syuzhet [20], developed at the time by the director of the Nebraska Literary Lab, Matthew Jockers, Associate Professor of English at the University of Nebraska, Lincoln.

The package makes use of lexicons that have word-sentiment and word-emotion associations. Syuzhet can invoke one of four such lexicons: bing [21, 22], afinn [23], stanford [24, 25] and nrc [26]. The algorithm identifies words from the posts with words connected with sentiment and emotions in the lexicon. As an explicit example, researchers at the the National Research Council Canada (NRC), for which the 'nrc' lexicon is named after, chose a list of 14,182 words and crowdsourced it out to 2000 volunteers to attribute sentiment and emotions to these words. They decided on eight standard emotions: anger, anticipation, disgust, fear, joy, sadness, surprise, and trust, of which the participants could select a single, combination or none of these emotions. This is not to say these are truly the eight basic emotions [27] but the researchers resolved on using Plutchik's selection of eight "basic" emotions [28], making a small adjustment to his list by replacing one of his emotions, acceptance, with trust. The choices of sentiment were either positive, negative, a combination of both or none. We tested each of these four lexicons and found no significant difference between any specific one. For that reason, we continued to work with the 'nrc' lexicon since it and the 'stanford' lexicon comes with a built-in function (**get_nrc_sentiment** or **get_stanford_sentiment**), for practicality. Figure 2.1 shows a basic diagram of how Syuzhet words in conjunction with the 'nrc' lexicon. The NRC HITS are words in the textual corpus that match with words in the 'nrc' lexicon. The next session gives a rundown of the code and more detail as to how the matrix output is generated.

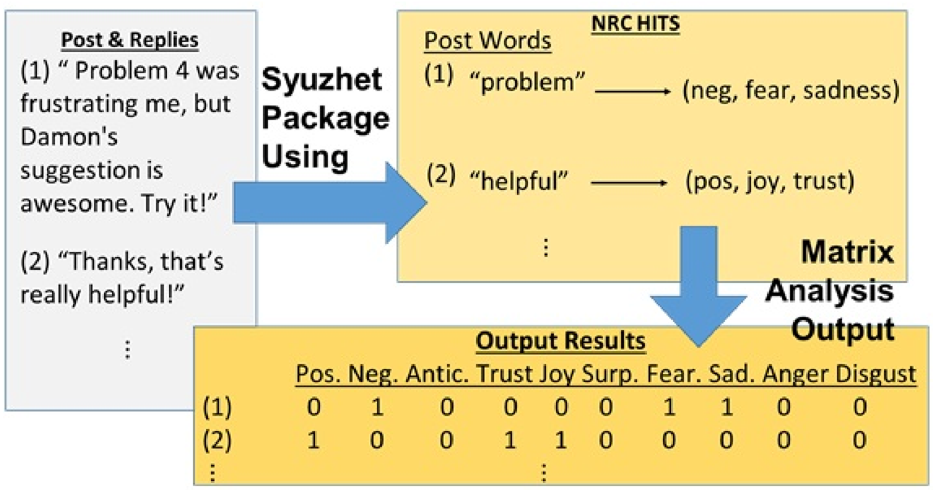


Figure 2.1: A Visualization of the Syuzhet Package Process

2.3 How SA Code Works

The way sentiment analysis works can be broken down into steps; the steps can be followed in the Code: Sentiment Analysis in the Appendix, if familiar with R code and syntax.

To help visualize the steps, we will use the bottom but not the first or last temporally made post, at position 2971, from the Fall 2015 semester data set:

studentdata[[2971,6]] → "All hail the engineers from all over! i'm pretty sure this class is going to be exciting."

Step 1: The first step is to import the semester data. We first arrange everything in Excel as an .xlsx file, with a data enclosed in separate file sheets titled by its semester year.

The headers to each column of information are titled student, score, gender, date, time, and text as illustrated by the below table containing our demonstrated post. There were more columns of information provided by CN such as the number of: pictures, other attachments, reflections and likes. But the direction of our studies implementing sentiment analysis made no use of these variables.

student	**score**	**gender**	**date**	**time**	**text**
⋮	⋮	⋮	⋮	⋮	⋮
77765428	93.37603	M	8/24/2015	10:14:13 PM	"All hail...exciting."

All the discussions and scores are imported as a data.frame in R via the **read.xlsx()** function from the package 'xlsx' into a variable called "studentdata". For sentiment analysis, we are only interested in the posts/replies, namely column 6 or within the data.frame header, $text; columns in R are described by "$" combined with the name of the imported column "text".

```
messages <- studentdata$text
```

Step 2: The next step is preprocessing the posts and replies, cleaning and preparing the data for the correct input structure used by the Syuzhet package. Each entire post is initially a single string from the data.frame. The posts can include many words and special characters that won't benefit our analysis. So we lowercase all words using **tolower()**, delete numbers, apostrophes, punctuations such as exclamation marks, questions marks, periods, hyphens, etc., and also the empty spaces before and after the end of the posts using the **gsub()** function. Then we can break up the posts into its individual word strings using **strsplit()**:

doc.list[[2720]] → "all" "hail" "the" "engineers" "from" "all" "over" "i" "m" "pretty" "sure" "this " "class" "is" "going" "to" "be" "exciting"

These are stored now as a list in R, which is like an array or vector but can hold uneven number of words in each row, where each row designates a complete post by a student.

Step 3: We created a stop words list (Appendix C), or words that don't convey much meaning such as "a" and "the", by finding freely available stop word lists online. We later created a masking word list (Appendix D) when we later found sentiment for words not in the correct context of the course such as "physics", "problem", and "force". We created a .txt file for all physics terminology supplied on the course syllabus, words that express a different meaning in reference to the physics course. In addition, R comes with a stopword package to further eliminate words that have no effect on SA. The function with the stopword list is called **stopwords("SMART")**, accessing the SMART (System for the Mechanical Analysis and Retrieval of Text) Information Retrieval System developed by Cornell University back in the 1960s. We appended all lists together, for rigor.

Another feature is to stem the word. Ideally this is to eliminate verbs and nouns of their "s" and "ing", thereby increasing the capacity to match words. Unfortunately, we haven't found a package to stem words in such a manner. A common package for stemming in R is called SnowballC. We found it to stem many words correctly such as "problems" to "problem" but this is not entirely consistent. For instance, the word "acquiring" gets stemmed down to "acquir" or "ability" to "abil"; the words "acquiring" and "ability" both have sentiment and emotion associated to them but not to these stemmed versions.

The program applies the stop words to eliminate those words from all the posts, or word corpus, via the **tm_map**() function from the package 'tm'.

The **tm_map()** function is inside a 'for loop', with changing index 'var', which goes through each word in each row and removes words that happen to be in the stopword list. Below shows what our example post looks like after this process, changing variable from doc.list to message:

message[[2720]][1] → "" "hail" "" "engineers" "" "" "" "" "" "pretty" "" "" "class" "" "" "" "" "exciting"

Step 4: That brings us to the main step in the code, applying sentiment analysis. The code in this step starts by initializing all the variables, used to collect and store the results from the analysis. Inside the 'for loop' whose purpose is to find all the word-lexicon matches for each post, we wanted to exclude all the blank double quotation marks. A method to simply remove empty double quotations "" in R doesn't seem to be available, so a way around this is to convert "" into *NA*. R does have a function called **na.omit()** that removes these pesky blanks. After preprocessing, applying stop words, and removing blanks, we are left with these filtered words.

The row from the variable message, signifying all the words in a single post, is stored accordingly in the a single variable after filtering:

w → "hail" "engineers" "pretty" "class" "exciting"

Supplied with just the filtered words of a single post saved as a list in the variable w, we can apply sentiment analysis using the function: **get_sentiment**(w, method="nrc") to find the sentiment per words in the post and **get_nrc_sentiment**(w) to find the corresponding emotions per words in the post.

The result of the sentiment analysis is a logical list of numbers, 1 or -1 signifying a positive or negative match, respectively, or 0 representing a neutral or no match:

		"hail"	"engineers"	"pretty"	"class"	"exciting"
sentiment_w	→	0	0	1	0	1

emotion_w →

	anger	ant.	disgust	fear	joy	sad.	surp.	trust	neg.	pos.
"hail"	0	0	0	0	1	0	0	1	1	1
"engineers"	0	0	0	0	0	0	0	0	0	0
"pretty"	0	1	0	0	1	0	0	1	0	1
"class"	0	0	0	0	0	0	0	0	0	0
"exciting"	0	1	0	0	1	0	1	0	0	1

Similarly, the emotional analysis produces a list of the eight emotions [anger, anticipation, disgust, fear, joy, sadness, surprise, and trust] as well as redundantly corresponding positive and negative matches, each row representing a word in sequence with the list of words in the post:

The reason we use the previously repetitive sentiment function is to incorporate the match listing as a logical index array to find the corresponding words in w:

wordpost_w <- w[as.logical(sentiment_w)] → "pretty" "exciting"

Interestingly, the word "hail" is excluded since it has both a positive AND negative connotation. The word is ignored due to its lack of overall input when summing up all sentimental contributions.

After saving all these variables to their corresponding lists, we also found out which words confer either a positive and negative association and saved these in the lists 'positivewordhit' and 'negwordhit', respectively. All the positive and negative words in these two lists are disarranged, the rows representing each post merged into one, with the function **unlist()** upon the completion of the 'for loop'. We examined this list to see common positive and negative words, which impacted our masking word list aforementioned, and we will cover this in more detail in the Results chapter.

The last part is counting sentiment for each post and counting all emotion in each emotional category. The effect is to aggregate all the sentiment and emotion to be linked with each single post. The resulting exemplary post will appear in the cumulative sentiment matrix for all posts:

sentiment_matrix →

totalsent
⋮
2

And correspondingly, the emotion matrix:

emotion_matrix →

anger	ant.	disgust	fear	joy	sad.	surp.	trust	neg.	pos.
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
0	2	0	0	2	0	1	2	1	3

Step 5: The last step is saving the results. We append the sentiment_matrix to the data set "studentdata" and then additionally, emotion_matrix to the data.frame "studentdata". The final step is writing the data to an Excel spreadsheet designated for results, by appending a sheet with the year of the semester to it. If the sheet is already present in the spreadsheet, the function **write.xlsx()** will overwrite the existing sheet.

2.3.1 Word Frequency Determination & Word Cloud Function

The code in Appendix B can be plugged directly into the end of the code in Appendix A. It assesses the positive, negative and all words stored in the 'poswordhit', 'negwordhit', and 'WORDCORPUS' variable lists, respectively. We use these to find the word frequency in each separate list. The word frequency gives us insight into the operation of our sentiment analysis by looking at the words that it appears to measure. We will discover some words that shouldn't be detected either because they were incorrectly parsed in the preprocessing section of Appendix A or improperly interpreted in context to the course, and some words that question the usage of one word as compared to another synonymous word.

Continuing from where we left off in Appendix A, we begin by calling upon two extra packages through the function **library()**. The packages are 'wordcloud' and 'rcolorbrewer'. The main package is 'wordcloud', which will produce a tag cloud or word cloud that represents individual words from the text corpus. The words impact will be depicted by size and color. The color selection stems from 'rcolorbrewer', a package in R that has pre-made color palettes to enliven plotted graphs such as word clouds.

The list of words, 'poswordhit', 'negwordhit', and 'WORDCORPUS', are converted into the proper data structure via **Corpus**(**VectorSource**()) using the package 'tm' as called upon in Appendix A. **VectorSource**() interprets each element in the vector x as a document. A vector is defined, in computer science terminology, as a collection of elements that are of the data modes: character, logical, integer or numeric. The entire vector, specifically all the elements, would be comprised of a single data mode. Our use of the data type, lists, in our code is technically a vector, therefore the list will be inputted into the **VectorSource**() function. The elements of the vector correspond to a student's post. Each post translates as a document, in computer science lingo. **TermDocumentMatrix**() creates a document-term matrix or term-document matrix, which is a mathematical matrix that describes the frequency of terms that occur in a collection of documents

Below is a demonstration on how documents, or student posts, are transcribed into a term-document matrix:

doc1 = "I am tired"
doc2 = "I am hungry" →

	I	**am**	**tired**	**hungry**
doc1	1	1	1	0
doc2	1	1	0	1

Although TermDocumentMatrix() produces a matrix, R doesn't interpret it as a matrix. **as.matrix**() converts the term-document matrix into an actual matrix data structure. The function **sort**() ranks the words, from their least frequent to most frequent word. Again we convert this data structure into another one, namely our matrix into a data frame through the function **data.frame**. The data frame can be stored permanently as a .csv file with the **write.csv**() function. Lastly a word cloud can be constructed with the **wordcloud**() function. The resulting word clouds for each semester is displayed at the end of Appendix B.

2.3.2 Determining way to categorize time of day students are active on CN

A last aspect to mention was an inherent difficulty we encountered in recording the cyclical nature of hourly time. We attempted to bin by hours the students' average CN time usage versus their course scores, but we confronted an issue with the wrapping effect of time. For instance if a student only posted at a time of 23:00 (11pm) and 1:00 (1am), then the average would be $(23+1)/2 = 24/2 = 12$. This would correspond to 12pm noon rather than midnight (24:00).

This difficulty was circumvented by binning students by their percentage of nightly activity, which we presented in the Results section. However as we initially attempted to compute an average of students' CN activity time, this would allow us to examine if there is any correlation with time of CN usage and students' average sentiment and overall course score. A possibility for this type of timely averaging may lie in a package in python or R, which we are looking into. Our expectation based on our current results side with very low correspondence between these factors. This is an ongoing problem that we are investigating to measure students' average CN usage time to their academic performance and average sentiment, to see if time of day based on their CN activity plays a role on student learning.

The Excel results spreadsheet now holds all the information about each posts' sentiment and emotion, as well as three additional .csv files with all words, positive word matches, and negative word matches. The values of the posts' sentiment and emotion are parallel in the rows of data, coupled to the individual students and the time and date of their posts, furthermore correlated with the students' score and gender. In the next section, we present the outcome into the investigation between these parameters and our coinciding sentiment analysis.

Chapter 3

Results

3.1 Preface

The ultimate goal of this work is to explore the value of data mining on our sets of data. Data mining offers itself as a useful tool, particularly on large or unwieldy amounts of data. We will see in some expected features, such as dips in positive sentiment and the positively ranked eight emotions at times of exams during specific weeks. But a real power of data mining is its ability to quantify affect through sentiment analysis (SA) and reveal details, swiftly and consistently. Moreover, the way we obtained our data is noninvasive. Unlike surveys, questionnaires and interviews, which have unintended sway on students answers [29], CN collects the students' posts in the natural setting of the course, in real-time as the student expresses their views and feelings on this social media platform.

We prepared much of our work for PERC Summer conferences, presenting figures taking advantage of the analysis done in R. Additionally, we investigated three research questions that we will formulate and outline in the following section. In this chapter we will present the exploratory results along with the outcomes from the research questions, ending with our current research focus of an upcoming conference proceeding paper.

3.2 Word Frequency

As described at the end of Chapter 2, the algorithm saves the aggregate sentiment of each post into a spreadsheet, appended to the overall data.frame of the imported data. But not all outcomes, resulting from the SA process performed by the code given in Appendix A, are permanently saved. The word matches are only stored as variables inside R, while R is running. In Appendix B, we supply additional code that makes use of these identified words, which give information on what individual words students chiefly used in their posts. The code saves the word frequency information as a spreadsheet and produces wordcloud images.

The word clouds helped illustrate common words, but the real insight came from viewing the most frequent words. Tables 3.1a, 3.1b and 3.1c tabulate the 10 most frequent words, 10 most positive words, and 10 most negative words for all semesters with the stop words and course maskwords applied. These are the same lineup of words found in the saved .csv files for most common, positive and negative words that the code in Appendix B generates.

It turns out that some of these words have a different emotional and sentimental association in the NRC lexicon than what is meant by in the course. The word “force” has a negative connotation since it takes the meaning to be violence upon someone or unwilling obligation to do something rather than the physical meaning of the change of momentum with respect to time. Hence we added the physics keywords and terminology from the course syllabus to the masking words and removed all the words to avoid skewing the data from the physical meaning rather than their social equivalent meaning. Another demonstration of unintended word connotation is the word ‘collision’, which turns up in Table 3.1c under Fall 2015. The word ‘collisions’ was pulled from the syllabus as a word from physics terminology, with no negative connotation. But forgetting to add the singular case of the word to the masking words, resulted in 14 times ‘collision’ appeared as a negative word in the semester. When adding ‘collision’ to the masking words, the word ‘hit’ takes its place occurring also 14 times under Fall 2015.

Equivalently, there are other physics and inadvertent words disguised as actual words that made word-lexicon matches. An example is the unexpected word, ‘sin’ that appears 32 times in the Fall 2016 semester. We looked for the word in context of transgressing an offense from the data corpus in all posts only to find students have never used this word in such a manner. Instead when parsing the data, the word ‘sin(x)’ as the sinusoidal function was separated from the parenthesis (‘sin’ ‘x’) with the ‘x’ removed by the SMART stop words. This left the word ‘sin’ which the ‘nrc’ lexicon attributed the negative sentiment and emotion. This case is shown in Table 3.1c and was added to the masking words.

Another similar case is the word ‘character’. No use of the word ‘character’ appeared in sync with the actual posts. Careful inspection of the analysis structure of SA revealed that an empty list element fills in as ‘character(0)’, meaning no characters or strings lie inside this list location. Like ‘sin(x)’, the parenthesis are stripped (‘character’ ‘0’) and the number ‘0’ is removed by the SMART stop words. We resolved this problem by parsing the posts better, inputting the list directly in the Syuzhet SA function rather than attempting to convert the list into a vector data structure which erroneously was believed to be the way to run the Syuzhet functions.

One last aspect of word frequency that we were intrigued by was the use of the word ‘exam’ as compared to the use of ‘test’. These are synonomous words with the same length of characters, so students shouldn’t select one more so than the other for simplicity of typing. But in Fall 2015, the difference in the word ‘test’ was more than twice that of ‘exam’. This is a segway into the discussion about topic modelling, which we will further cover in Chapter 4. Topic modelling may also shed light on words that aren’t meant in the proper context. The word ‘fall’ in Table 3.1c for Fall 2014, 2015 and 2016, may not refer to the act of declining or admitting defeat or failure but instead, is a physics word for objects subjected to gravity or rather the “fall” semester since it shows up primarily in the fall semesters.

Table 3.1: Most Frequent Words: Common, Positive, and Negative

(a) 10 Most Common Words & their Frequency

Fall 2014		Fall 2015		Fall 2016		Spring 2017	
word	**freq**	**word**	**freq**	**word**	**freq**	**word**	**freq**
time	193	test	423	good	343	good	387
good	156	time	279	test	261	time	373
exam	155	good	270	time	256	test	347
class	149	study	236	lab	243	exam	324
ball	124	class	229	exam	207	homework	268
study	123	answer	205	practice	178	class	260
homework	112	question	194	great	175	hope	254
mass	112	exam	190	homework	174	feel	235
test	103	homework	158	video	166	seeds	221
work	98	semester	152	class	161	question	193

(b) 10 Most Positive Words & their Frequency

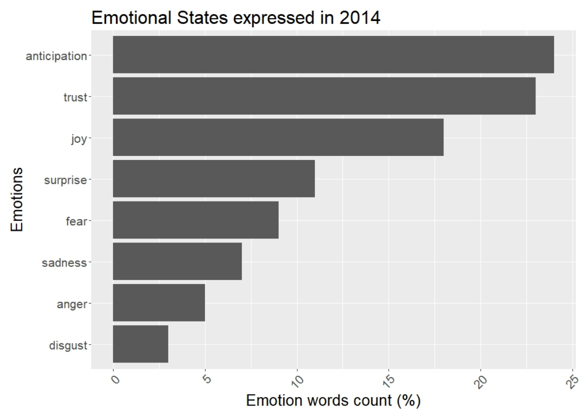
Fall 2014		Fall 2015		Fall 2016		Spring 2017	
word	**freq**	**word**	**freq**	**word**	**freq**	**word**	**freq**
good	156	good	270	good	343	good	387
study	123	study	236	practice	178	hope	254
question	94	question	194	helpful	148	question	193
hope	92	practice	139	hope	139	study	191
pretty	86	hope	139	question	139	practice	187
center	72	luck	124	study	126	luck	173
practice	70	helpful	99	luck	121	pretty	170
luck	68	extra	86	pretty	97	extra	111
major	68	pretty	72	love	69	agree	110
extra	68	major	62	working	66	interesting	105

(c) 10 Most Negative Words & their Frequency

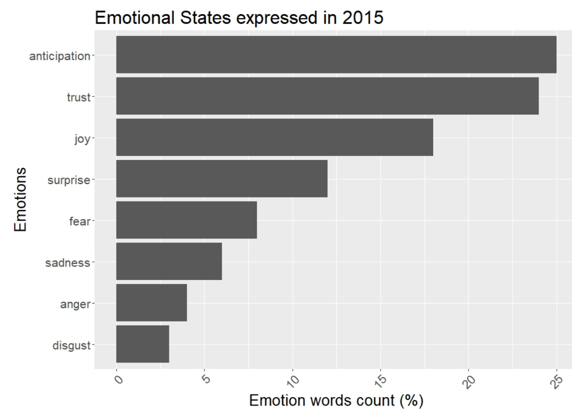
Fall 2014		Fall 2015		Fall 2016		Spring 2017	
word	**freq**	**word**	**freq**	**word**	**freq**	**word**	**freq**
wrong	37	wrong	61	wrong	35	forget	68
fall	36	bad	30	sin	32	bad	57
wait	34	forget	29	fall	29	wrong	53
small	23	fall	25	case	27	wait	44
bad	23	late	24	forget	25	nervous	35
case	19	nervous	24	hit	24	worried	33
forget	17	wait	22	worried	19	case	29
collision	17	small	16	spent	15	late	26
leave	14	case	15	wait	15	lost	21
falling	14	collision	14	bad	15	hit	21

3.3 Sentiment Analysis

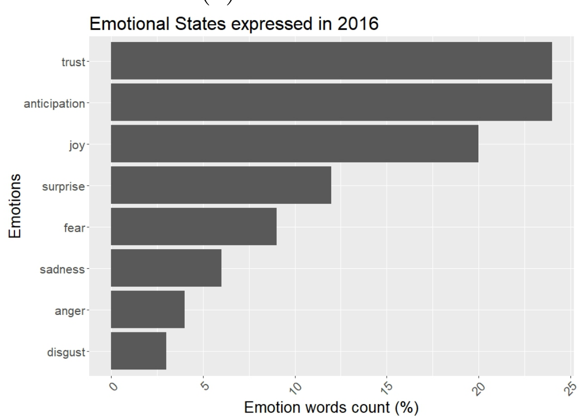
The distribution of words gives us an initial intuition into the words that were extracted by sentiment analysis. The next step is to make use of the quantification of the sentiment/emotion values of each post. Their scores, gender, and CN activity is directly connected to the student and each of their posts.



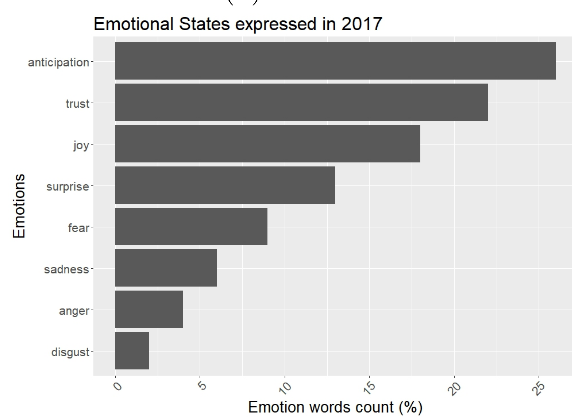
(a) Fall 2014



(b) Fall 2015



(c) Fall 2016



(d) Spring 2017

Figure 3.1: Semester percent distribution of 8 emotions

We directed our first attention to the overall semester. The power of this tool, namely an automated method to code and quantify words, is in the abundance of words. The whole semester contains roughly 3000 posts as exhibited in Table 1.2, each comprised of a multitude of words which had a myriad of matches to the 'nrc' lexicon. We wanted to capture the emotion for an entire semester. Figure 3.1 lays out the emotional distribution for the four semesters. The arrangement of emotions from largest to smallest tends to be in the order of: anticipation, trust, joy, surprise, fear, sadness, anger, and disgust. Only in Fall 2016 does trust barely replace anticipation for highest emotion over the entire semester. We chose to rank the emotions from most positive connotation: joy, trust, anticipation, surprise, sadness, fear, anger, and disgust. This is our own interpretation of the emotional overtone of the eight available emotions. The comparison between the semesters and our own ranking shows mostly positive emotions, the four most positive emotions appearing out of order but on top, and the four more negative emotions appearing also disordered but towards the bottom of the semester percent distributions. The distributions satisfies the impression observed in the classroom, at least by the standards of the intention of having a largely positive class chat room and by the limit of this tool to draw out a quantitative value on the sentiment from the course's online discussions.

Moving on to the data stored in 'studentcndataresults.xlsx' at the end of the code in Appendix A, most of the munging of the data happened in Microsoft Excel. We convert the total number of sentiment and emotional matches into percentages per post. Remember the sentiment and emotion was a matrix of the total sentiment/emotion-word matches per post. We first calculate the average sentiment for each student. Then we comprehensively average student sentiment for each semester or combination of all semesters from this. Figure 3.2 features the positive and negative sentiment and the eight emotions stacked for all four semesters. The error bars between the sentiment are standard error (SE) of the mean defined by the standard deviation divided by the square root of the number of students per gender ($SE = \frac{\sigma}{\sqrt{n}}$). We will use this for all error bars in the ensuing graphs. We resort to present only sentiment differences in the subsequent results since it is simpler to compute and illustrate this binary differences rather than among eight different emotions.

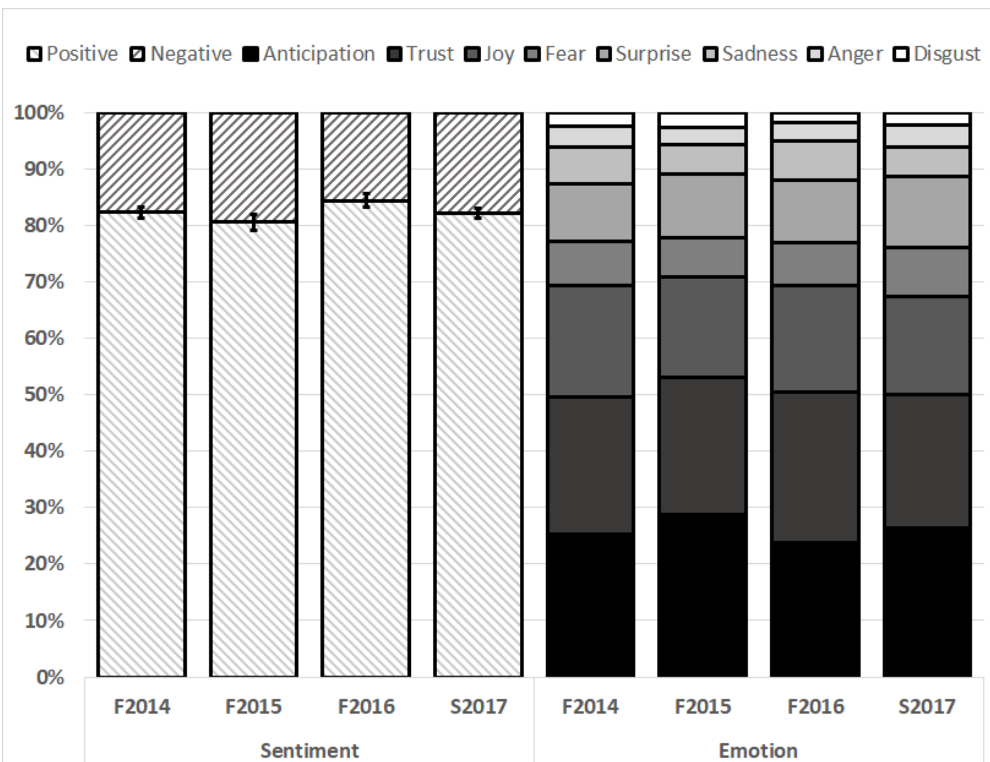


Figure 3.2: Percentage of stacked sentiment and emotion for the entire semesters of Fall 2014, Fall 2015, Fall 2016, and Spring 2017

Figure 3.3 illustrates the stacked sentiment with standard error bars and emotion per week, starting with the first week of the semester which corresponds to August 19th, ending in December 21st. We group the first two weeks and last two weeks into one week columns due to the low usage of CN at the start and end of the semester.

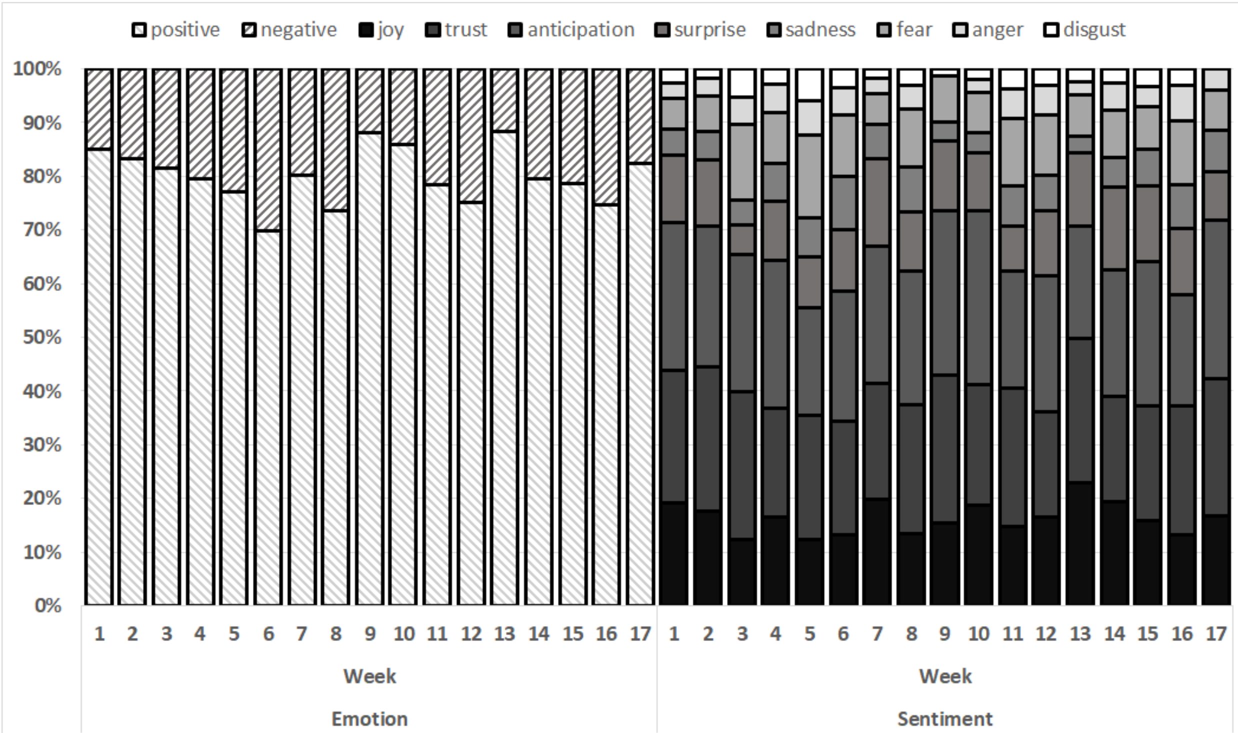


Figure 3.3: Percentage of stacked sentiment and emotion for each week of Fall 2015

Another way we examined the data is by week. The distribution by week reveals some interesting overlap between the happenings in the course. The dips in the positive sentiment and rise in negative emotions coincide with the three times that exams occur throughout the course. The idea behind a week by week sentiment/emotion evaluation of the course is a pursuit that CN could incorporate into an instructor feature to monitor student affect.

We received insightful research questions from reviewers of the PERC 2016 conference that we pursued and present in the following subsections. The first of those questions that we refined to acknowledged our prior findings was 'Can we use text mining methods to gain insight into class sentiment?'. We structured the questions that followed from the reviewers:

Sentiment vs Gender (3.3.1) Are there gender differences in the observed sentiment? Does course score reflect differences between gender?

Sentiment vs Activity (3.3.2) Are there differences in sentiment between students who participate heavily and those who do not? Are these differences also reflected by course score?

Another variable not tapped into is the time, recorded as a time stamp by the servers from when the post is received. We found that the number of students who use CN show no statistical difference in sentiment between those in the bracket of the most actively engaged and those less engaged (in subsection 3.3.2). Instead of a measure of engagement, we pursued the following exploration of time on student sentiment and performance. This study is still in progress but the current results will be presented in the last subsection. The research question raised is:

Sentiment vs Time of Day (3.3.3) Do students that post more at night express different sentiment than students that post only during the day? And again, does course score reflect this difference?

The next subsections will answer those research questions in the sequence that we examined them. We begin by looking at gender, if males or females show a difference in sentiment. Then we looked at sentiment comparison between students in bins of different frequency of posts, a gauge of student activity. The ensuing and last investigation is into the time of posts versus the sentiment and academic performance, namely students' course score.

3.3.1 Sentiment vs Gender

The data for each encoded post came with a binary gender assignment to each corresponding student. The first thing to mention is that the binary gender, male or female, is not as complete a set of descriptors as we might hope fore [30]. The IUPUI student directory only allows students to select from these two genders, but of course this excludes the identification of students belonging to transgender, genderqueer or non-binary. Working with this first order model, we wanted to see if there was a noticeable difference between men and women on CN.

We first had to find out the unique students from the data set, those students that used CN in the course. Excel can remove duplicates, so the column of all students by post can be diminished to just a singular case of the student code number. Recall that in order to keep with IRB (Institutional Review Board) policy, we had to encode student names for anonymity. The number of posts for each individual student was counted and the students' binary gender was identified. Additionally, the score was matched with each student that participated on CN, by making at least one post or reply. We arranged the data based on the students' gender. It was straight to follow that we computed the average sentiment, average course score, and average number of posts by gender for all respective students that used CN. Figure 3.5 depicts the average number of posts by gender as well as the average sentiment and average course score by gender for each individual semester. In order to keep analysis consistent, we removed those students that although they provided a post, that single or even multiple posts may not have registered on the 'nrc' lexicon with a sentiment. Those students with zero sentiment or who withdrew from the course, and thus had no course score, were taken out of the analysis. We held these student removals the same for all research question analysis.

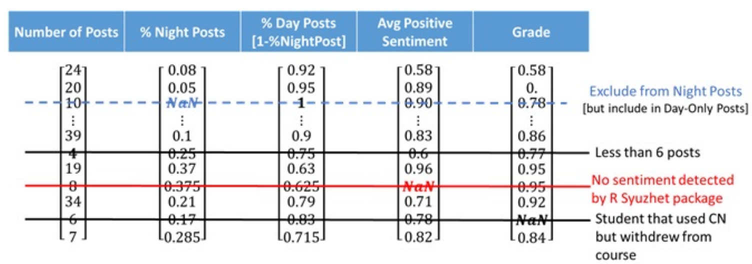
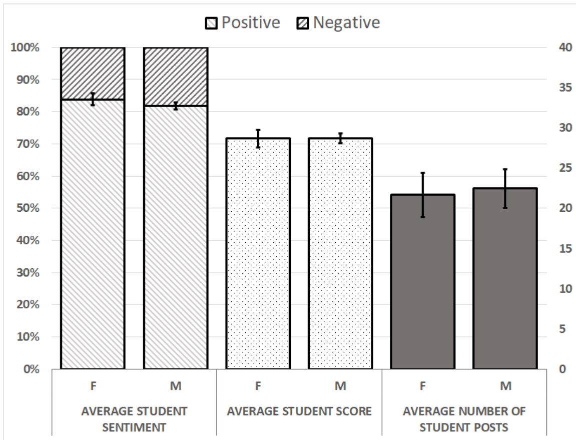
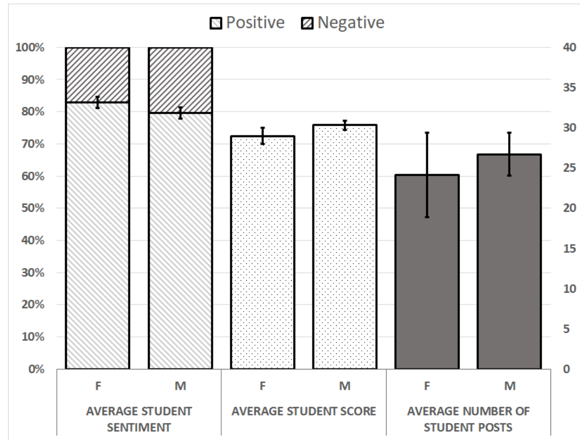


Figure 3.4: Data Selection: exclude withdrawals, no sentiment detection, less than a certain number of posts, and distinguish day-only CN users from nightly ones.

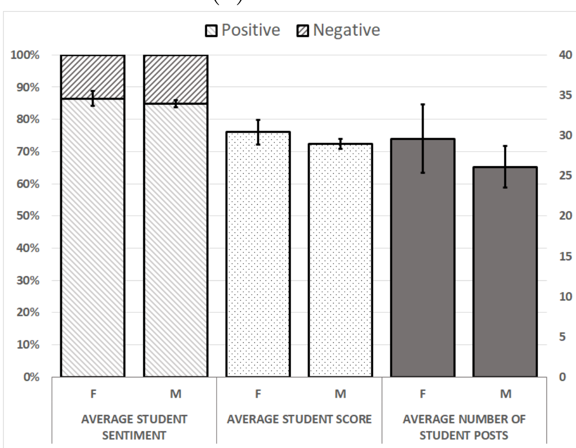
In addition, we looked at excluding students' that posted less than a certain number of posts. This shows up in the last analysis, the time of day activity. The data contains



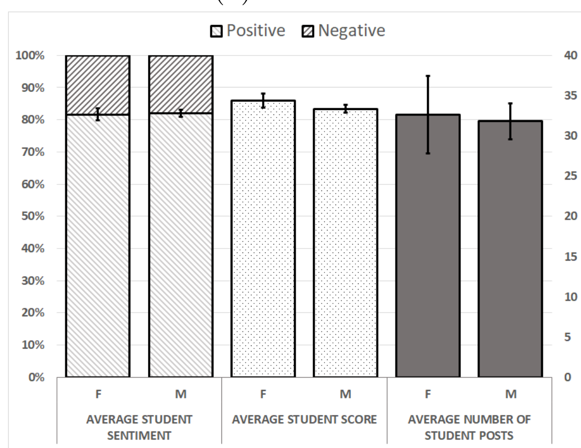
(a) Fall 2014



(b) Fall 2015



(c) Fall 2016



(d) Spring 2017

Figure 3.5: Female and male sentiment, course score, and average number of posts made for Fall 2014, 2015, 2016 and Spring 2017. The sentiment and score are measured on the left percentage axis and the average number of student posts is measured on the right axis.

the individual number of posts a student contributed in the course. It is therefore possible to restrain our analysis to students that contribute to CN more than a certain amount, such as 6 or more posts. This number comes from binning the students by their overall number of posts into four quartiles; six posts is the roughly the upper number of posts that the lower quartile of CN active students typically contribute. Figure 3.4 demonstrates the method to omit data points based on criteria; namely that students should provide more than a specific number of posts throughout the semester, their posts have ascribed sentiment values, the student concludes the course with a final score, and binning of night-time activity by percentage.

Since gender is the same throughout all semesters, we aggregated all semesters to create a larger effect size and looked at the relationship between these two genders. The result is depicted in Figure 3.6.

We ran a two-tail t-test between males and females and discovered the statistical p-values were too high to determine any statistical difference. That is to say, there was no statistical significance between either the average number of posts by gender nor average sentiment by gender. As a measure of certainty, we also compared the course score by gender and found no statistical difference there either. This helps to compare the affect, via the course sentiment, to the course performance, as roughly realized by the course score. The similarity between these two groupings gives a better sense of the validity of the sentiment analysis.

Fall 2014, Fall 2015, Fall 2016, Spring 2017 and the cumulative semesters exhibited a t-test p-value of 0.33, 0.28, 0.55, 0.89, 0.18, respectively for average sentiment. Similarly, the p-value for the statistical difference between course score for each semester did not dip below 0.2 for any semester. With closer examination, we noticed that with a smaller fraction of women to men in Fall 2016, Spring 2017 and cumulative semesters (Females/Males: 0.18 for 2016 and 2017, 0.24 for cumulative, 0.3 for 2014 and 2015), the sentiment looked less discernible with higher p-values between sentiments. To see the effect size, we calculated the Cohen D value for each semester and the cumulative semesters to notice than no semester exceeded a value of 0.2, signifying little importance. The average number of posts per student of a gender held consistently no statistical significance between males and females for all semesters. And the course score similarly had p-values greater than 0.2 for all semesters.

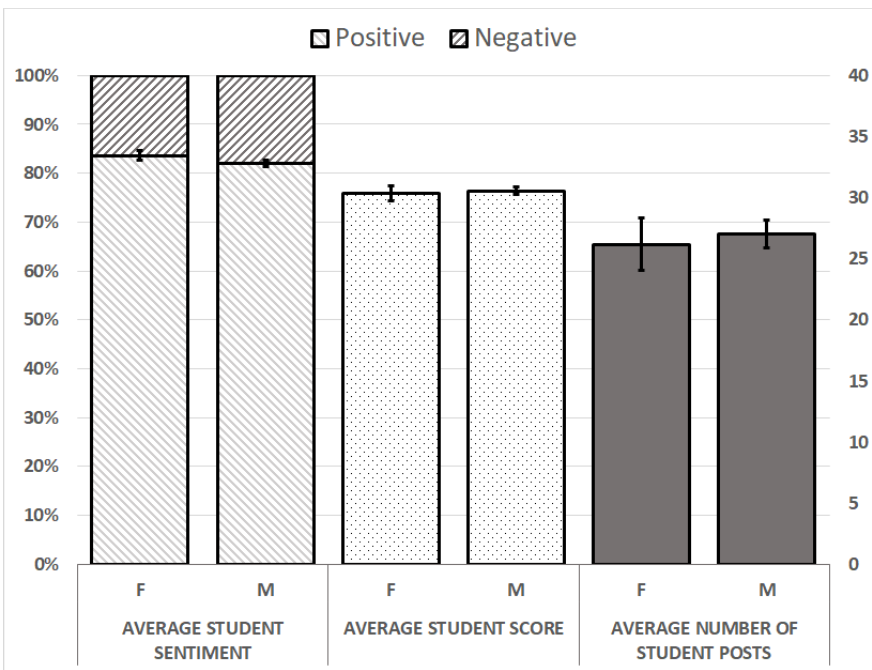
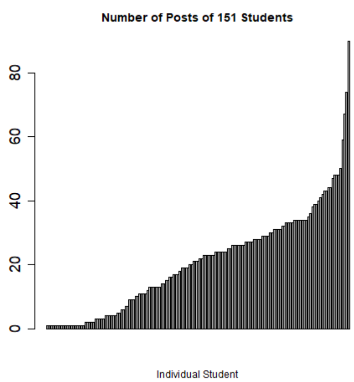


Figure 3.6: Female and male sentiment, course score, and average number of posts, cumulative over all semesters. The sentiment and score are measured on the left percentage axis and the average number of student posts is measured on the right axis.

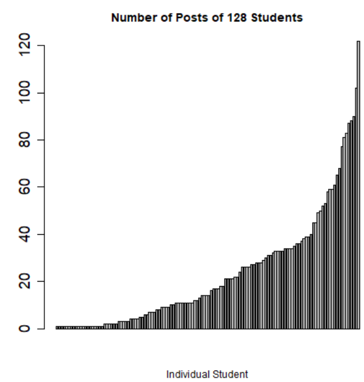
3.3.2 Sentiment vs Activity

Since gender proved to have no recognizable difference in sentiment or number of posts, we next looked at activity on CN. The idea that the most frequent users of CN would dominate the sentimental landscape was brought to our attention. We wanted to weigh the user activity, evaluated by the number of posts, of each student and contrast their sentiment with those students that posted fewer times, thereby justifying whether or not the sentiment happens to be skewed by these habitual CN users.

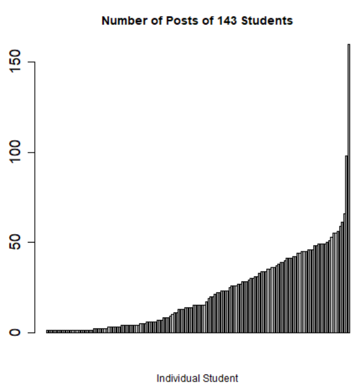
We approached this comparison by ranking each student by the number of their posts, regardless of gender, into four quartiles and examining the summed sentiment for each quartile. There has been a positive correlation discovered for CN usage measured by anar seeds and students' homework scores by Webassign [31]. We wanted to similarly see if the most active students showed a sentiment difference with more frequent students who posted less in four equal quarters of the number of post frequencies.



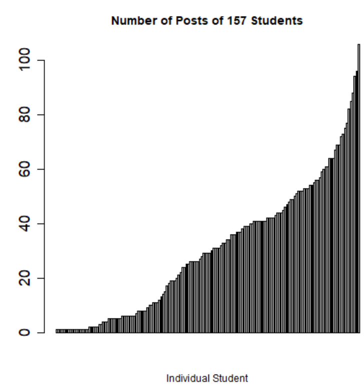
(a) Fall 2014



(b) Fall 2015



(c) Fall 2016



(d) Spring 2017

Figure 3.7: Number of posts made by each student in the semester

Instead of ranking the number of posts made by each students as displayed by Figure 3.7, we flipped the dependent variable and sorted the number of students by the number of ascending posts. The distribution now follows a frequency distribution for students who make a certain number of posts per the semester. Figure 3.8 gives a pictorial demonstration of how we split the students and their sentiment into quartiles by counting the number of students in increasing post frequency in the Fall 2015 semester. Figure 3.8, the visual representation of the quartiles, also depicts the sentiment per student (x), counted when falling into four positive percentage bins: either $0 \leq x \leq 25\%$, $25\% < x \leq 50\%$, $50\% < x \leq 75\%$, or $75\% < x \leq 100\%$. The results were stacked for each number of posts.

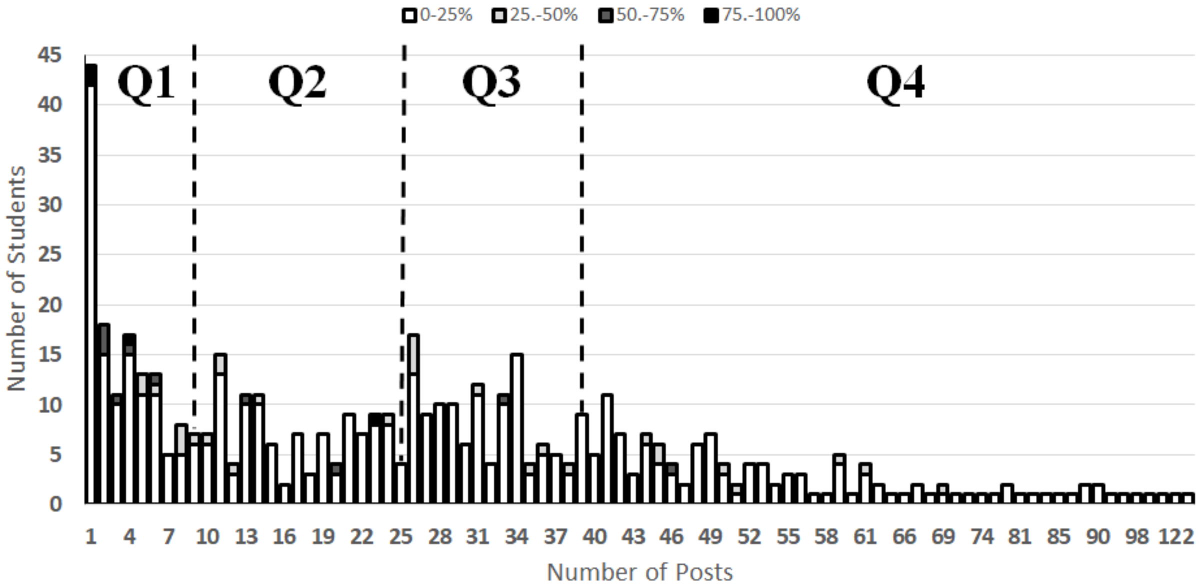


Figure 3.8: Number of students to the number of posts submitted for all cumulative semester. The students' sentiment is grouped into four increasing positive percentage bins.

We average the mean sentiment of each student that falls under the range of the frequency of posts. We evaluated the ranges by dividing the total number of students into four equal clusters. With the cumulative semesters, there were 566 students that possessed a sentiment post match, so each quartile has 124 students, the last two quartile having only 123 students. Corresponding to Figure 3.8, Quartile 1 (Q1) ranges from 1-8 posts containing 124 students, Quartile 2 (Q2) ranges from 8-25 posts with 124 students, Quartile 3 (Q3) ranges from 25-39 posts with 124 students, and Quartile 4 (Q4) ranges from 39-160 posts comprised of 123 students. Figure 3.9 has the mean positive and conversely the negative sentiment stacked for each quartile of students by activity.

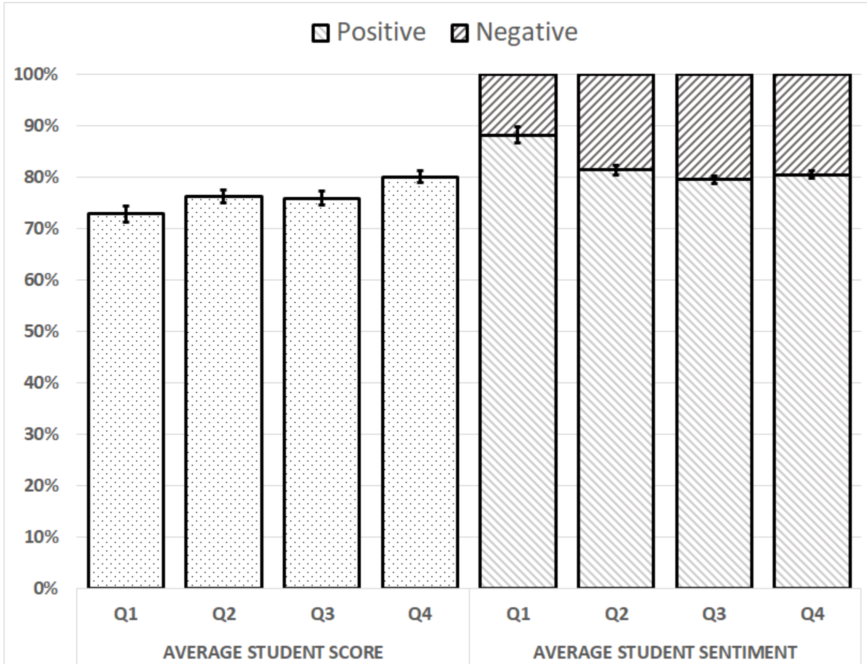
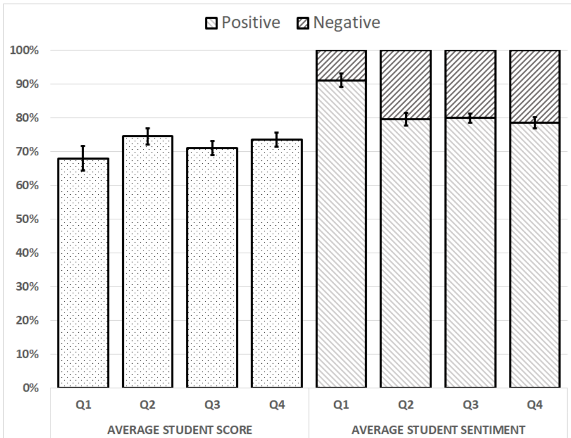
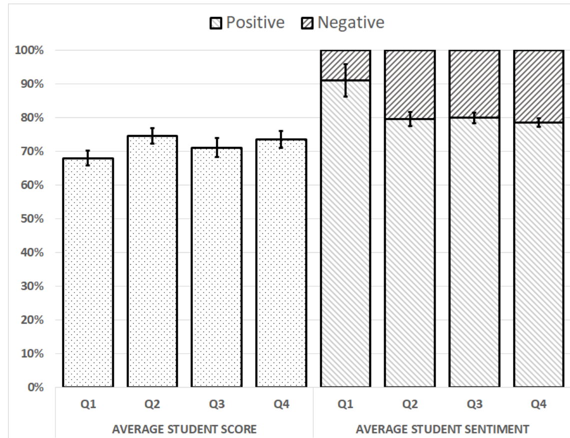


Figure 3.9: Quartile of frequent posts by CN users, cumulative over all semesters

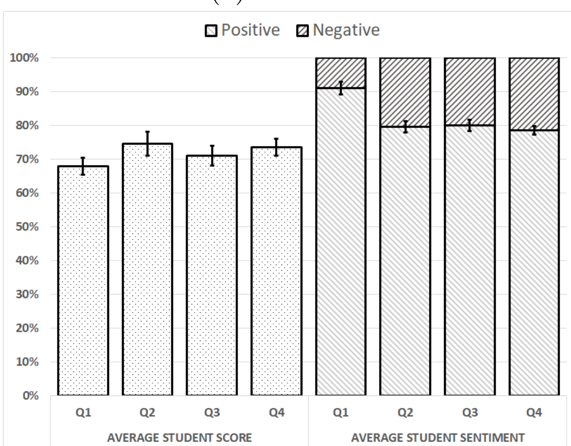
We performed a Tukey Significance Difference Test [32] to compare each of the quartiles. We found that the only significant difference was between Q1 and the other quartiles, which we attribute as possibly stemming from self-introductions encouraged at the first week of the semester.



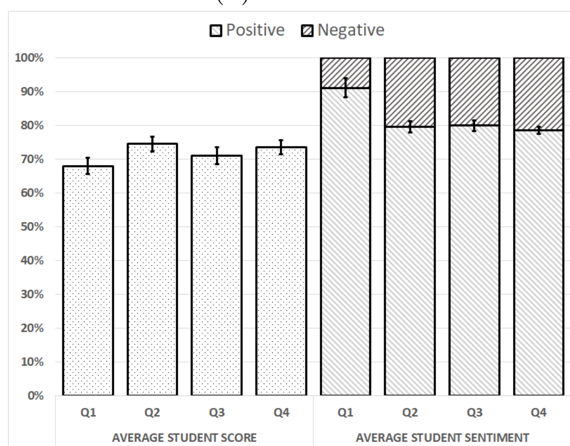
(a) Fall 2014



(b) Fall 2015



(c) Fall 2016



(d) Spring 2017

Figure 3.10: Quartile of frequent posts by CN users for Fall 2014, 2015, 2016 and Spring 2017 semester.

Figure 3.10 shows the quartile for each semester individually. The number of students that fall in each quartile ranges from 31-39 students. The Tukey Significance Test showed a similar finding for each semester, with a significant difference predominantly appearing for Q1.

3.3.3 Sentiment vs Time of Day

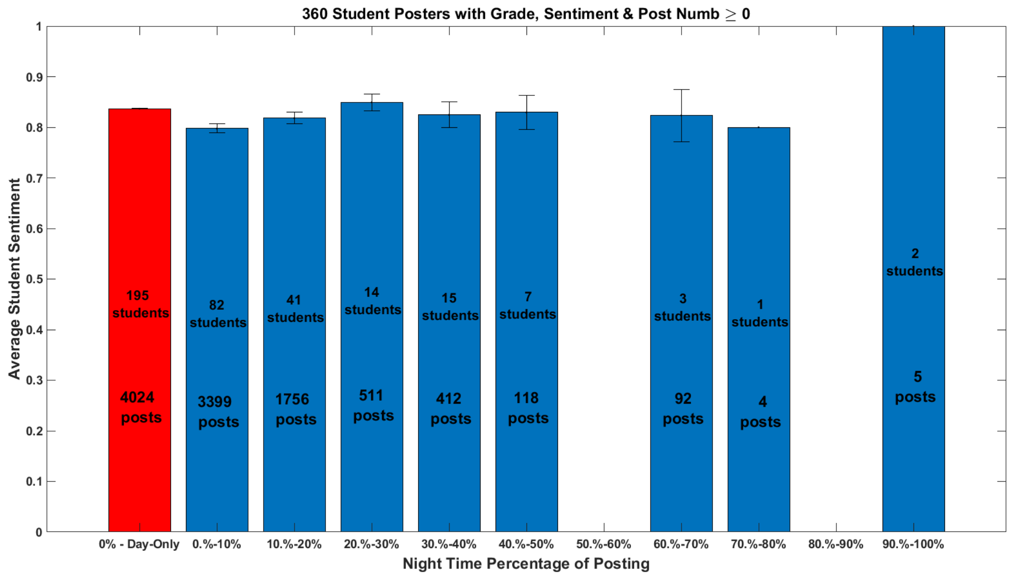
After discovering that neither gender nor the number of posts has any persisting statistical significant differences, we wanted to look at another variable, the time. There is a notion that more consistent students who do not stay up late should academically outperform late-night students. CN cannot monitor students' sleeping behavior but students who post at later hours can be ascertained by their posts' time stamp.

We figured out a time frame for late-night activity, and decided on the time between 11pm-6am. The plots of Figure 3.13 revealing semester activity in hourly time bins provided intuition as to when students are probably asleep by fewer CN activity. These are the students that we want to examine and compare to the students that post more during the day.

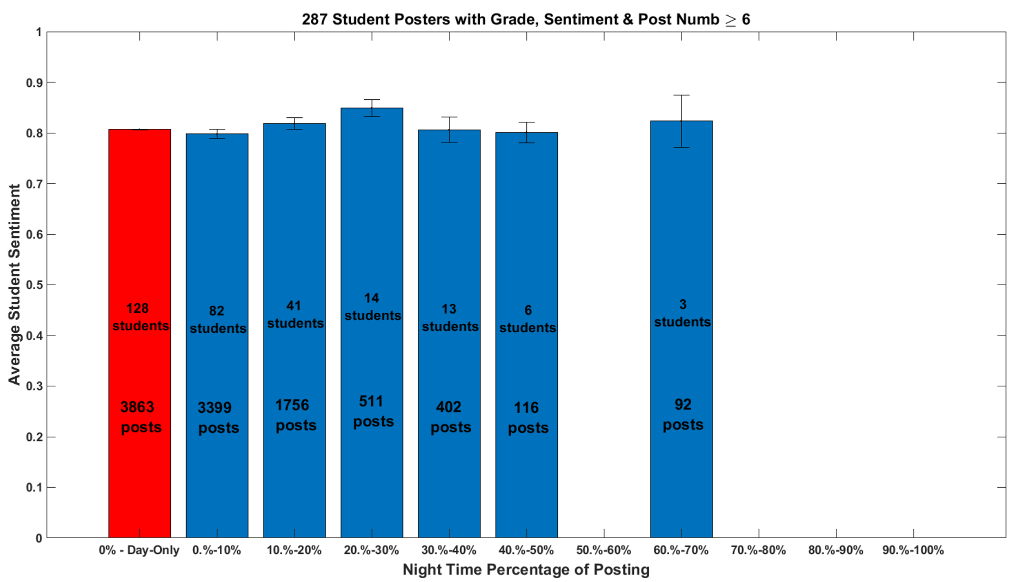
We returned back to the data and extracted the students that display some night time posts. Their scores will again be our measure of their overall academic performance. We removed those students that withdrew from the course and calculated their percentage of night-time activity, defined by their posts falling between 11pm and 6am. The first row of Figures 3.14, 3.15, 3.16, 3.17 shows all the students that made a night-time post, completed the course with a final score, and manifested a sentiment value from their posts. The last rows of Figures 3.14, 3.15, 3.16, 3.17 exhibits the same information as the plot with $n \geq 0$ in Figures 3.11 and 3.12.

Our goal is to compare the different percentage of night-time CN users to students that demonstrated no night time activity by using CN only during the day, never between 11pm≤post≤6am. To increase our sample size, we combined Fall 2015, 2016 and Spring 2017 together since they share the same course materials as illustrated in Table 1.2. We exclude Fall 2014 because that semester still used SmartPhysics, now called FlipItPhyiscs, before the transition to Webassign and Tipler for the subsequent semesters. Figures 3.11 and 3.12 shows the cumulative distributions for each week for all posts made by students who completed the semester with a course score and impressed an overall sentiment in the analysis. Additionally, the number of students and the number of posts are on display inside each bin.

The purpose for adapting our data allows us to return to the previous research question, whether frequency of posts has a role to play in the study. We considered this by investigating the consequence of selecting a threshold value of six posts, by removing most if not all students in Quartile 1 who showed the most statistical significance due to the abundant single self-introductory posts. The bottom graph in Figures 3.11 and 3.12 reveal the results; the similarity with it's adjacent figure visible by eye. We ran a Tukey Significance Difference test to uncover no statistical significance between any of the percentage groups.

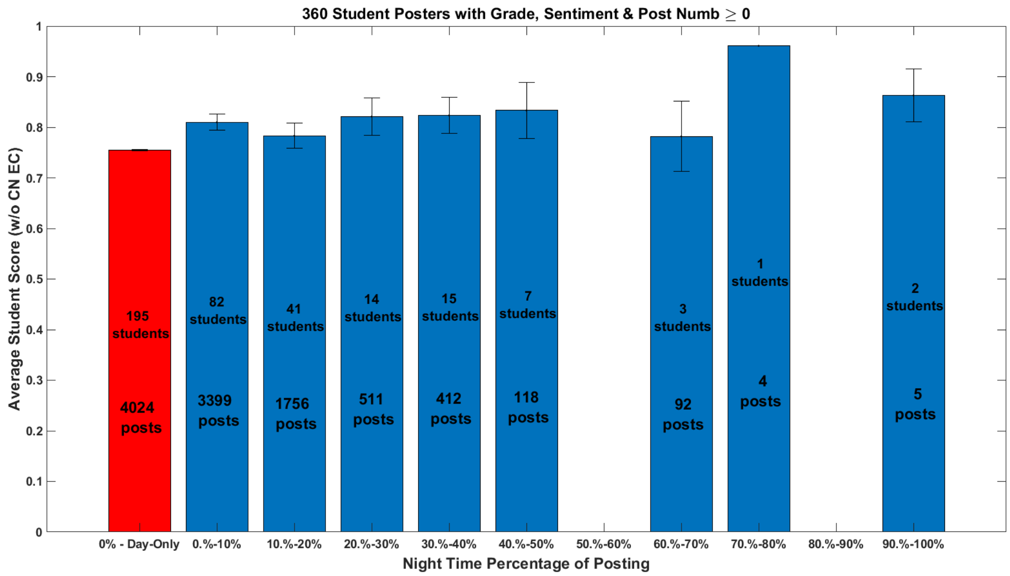


(a) Average sentiment percentage in each of the night time activity percentage ranges for Fall 2015, 2016 and Spring 2017 with zero or more posts

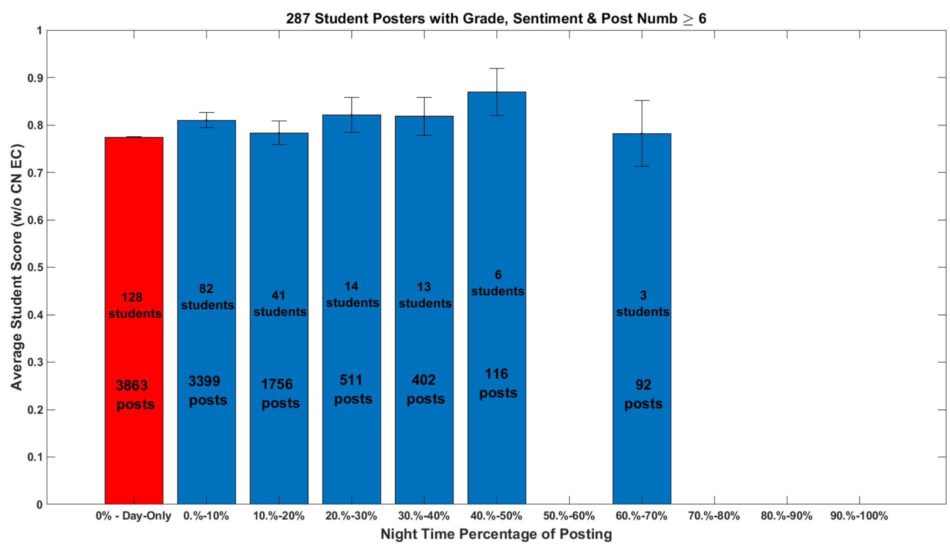


(b) Average sentiment percentage in each of the night time activity percentage ranges for Fall 2015, 2016 and Spring 2017 with six or more posts

Figure 3.11

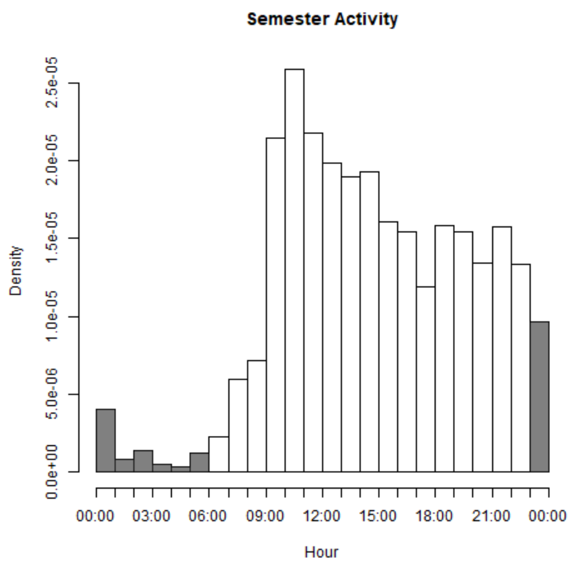


(a) Average course score in each of the night time activity percentage ranges for Fall 2015, 2016 and Spring 2017 with zero or more posts

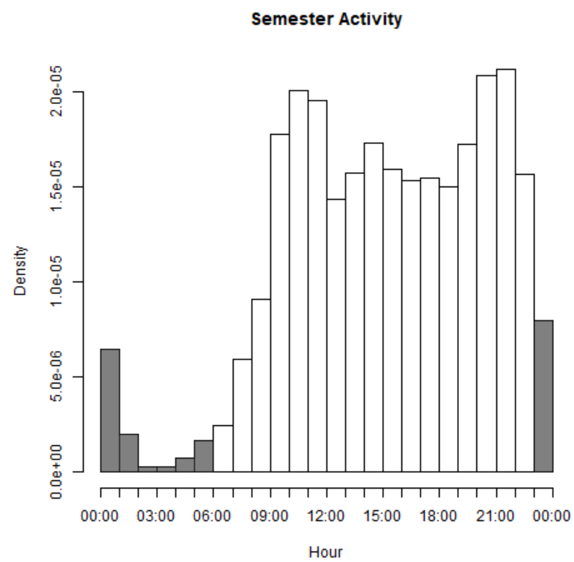


(b) Average course score in each of the night time activity percentage range for Fall 2015, 2016 and Spring 2017 with six or more posts

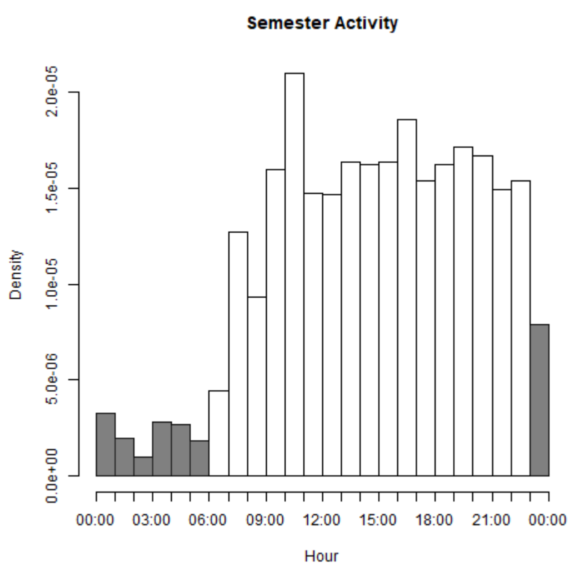
Figure 3.12



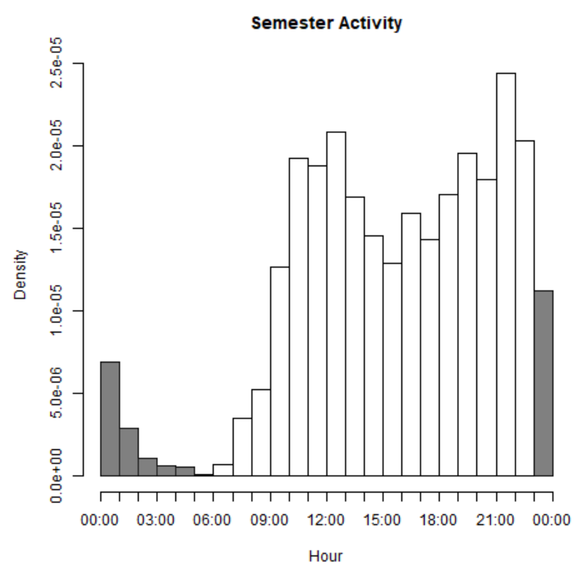
(a) Fall 2014



(b) Fall 2015



(c) Fall 2016



(d) Fall 2017

Figure 3.13: Semester density distribution of posts/replies made in hourly time-slots [grey-shaded night time postings between 11pm-6am]

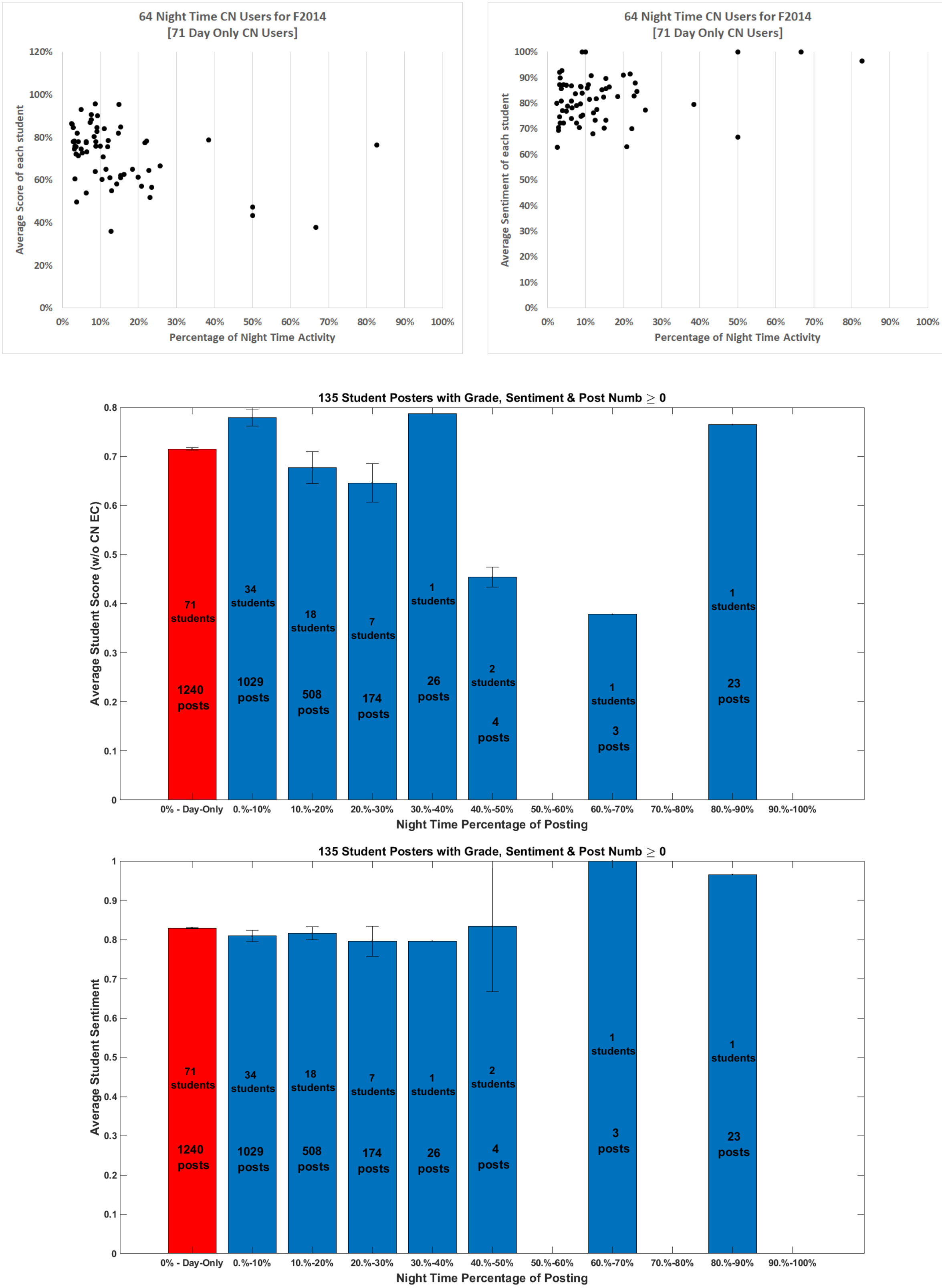


Figure 3.14: Average course score and sentiment scatter plot and bargraphs for each students versus their percentage of night time activity in Fall 2014

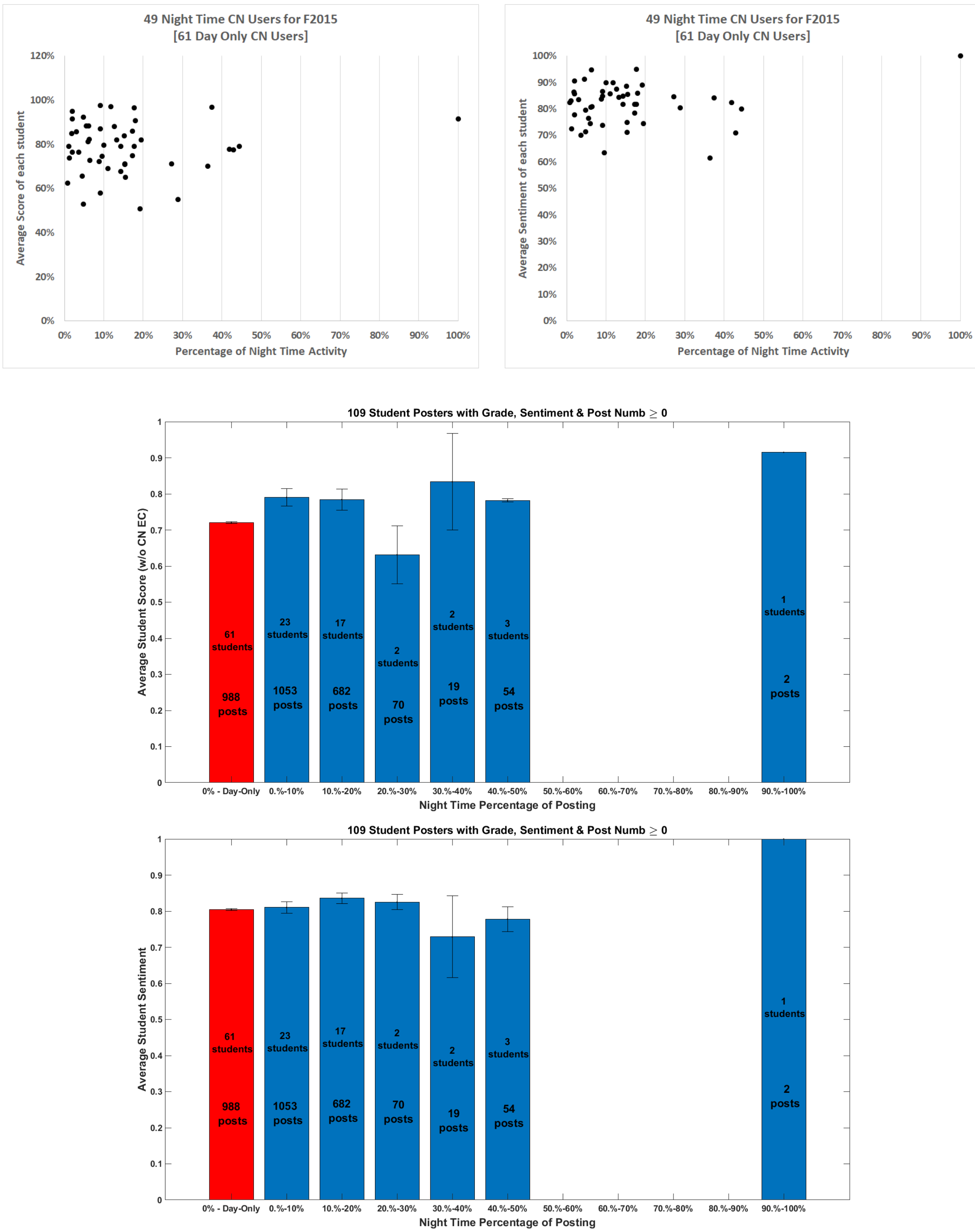


Figure 3.15: Average course score and sentiment scatter plot and bargraphs for each students versus their percentage of night time activity in Fall 2015

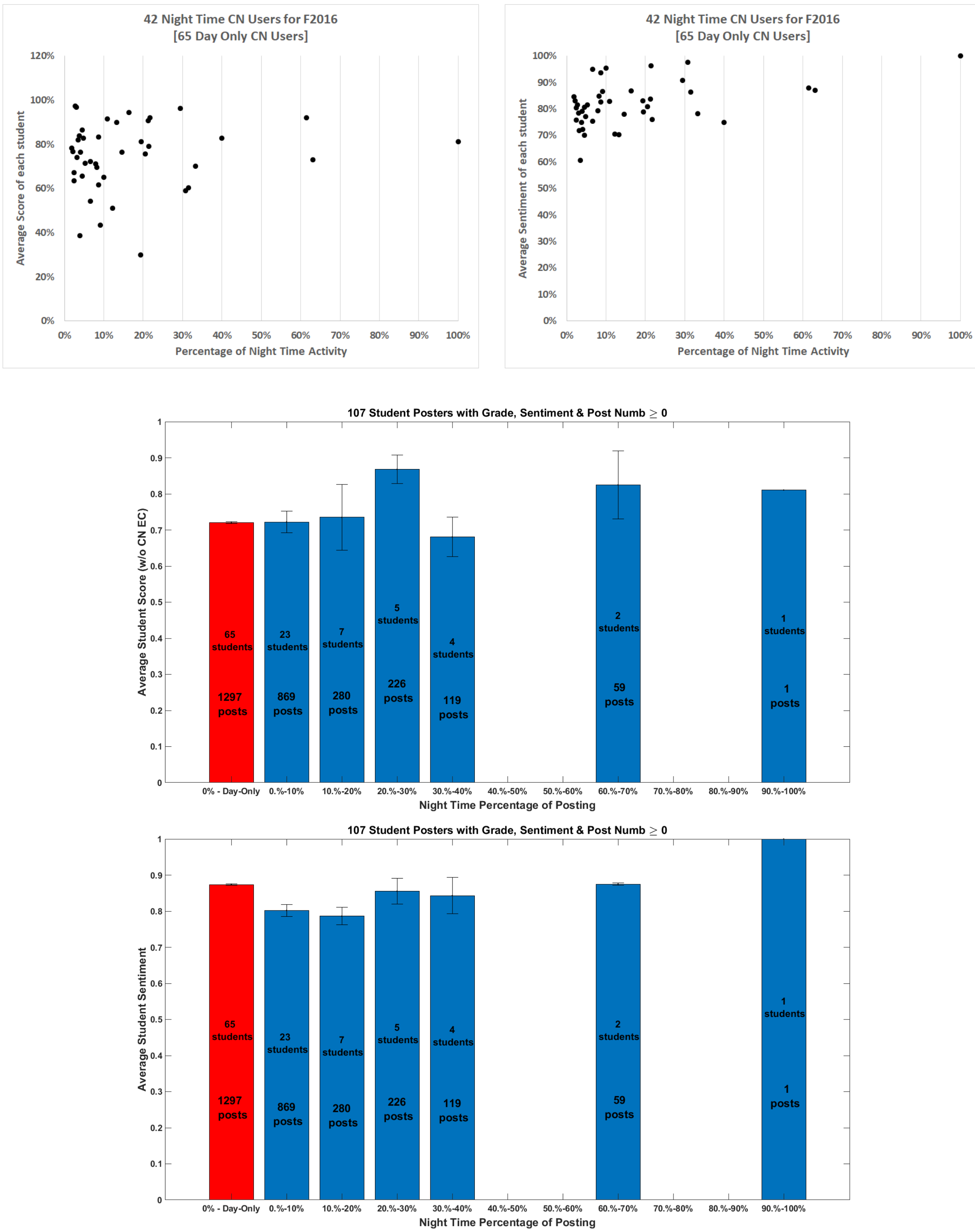


Figure 3.16: Average course score and sentiment scatter plot and bargraphs for each students versus their percentage of night time activity in Fall 2016

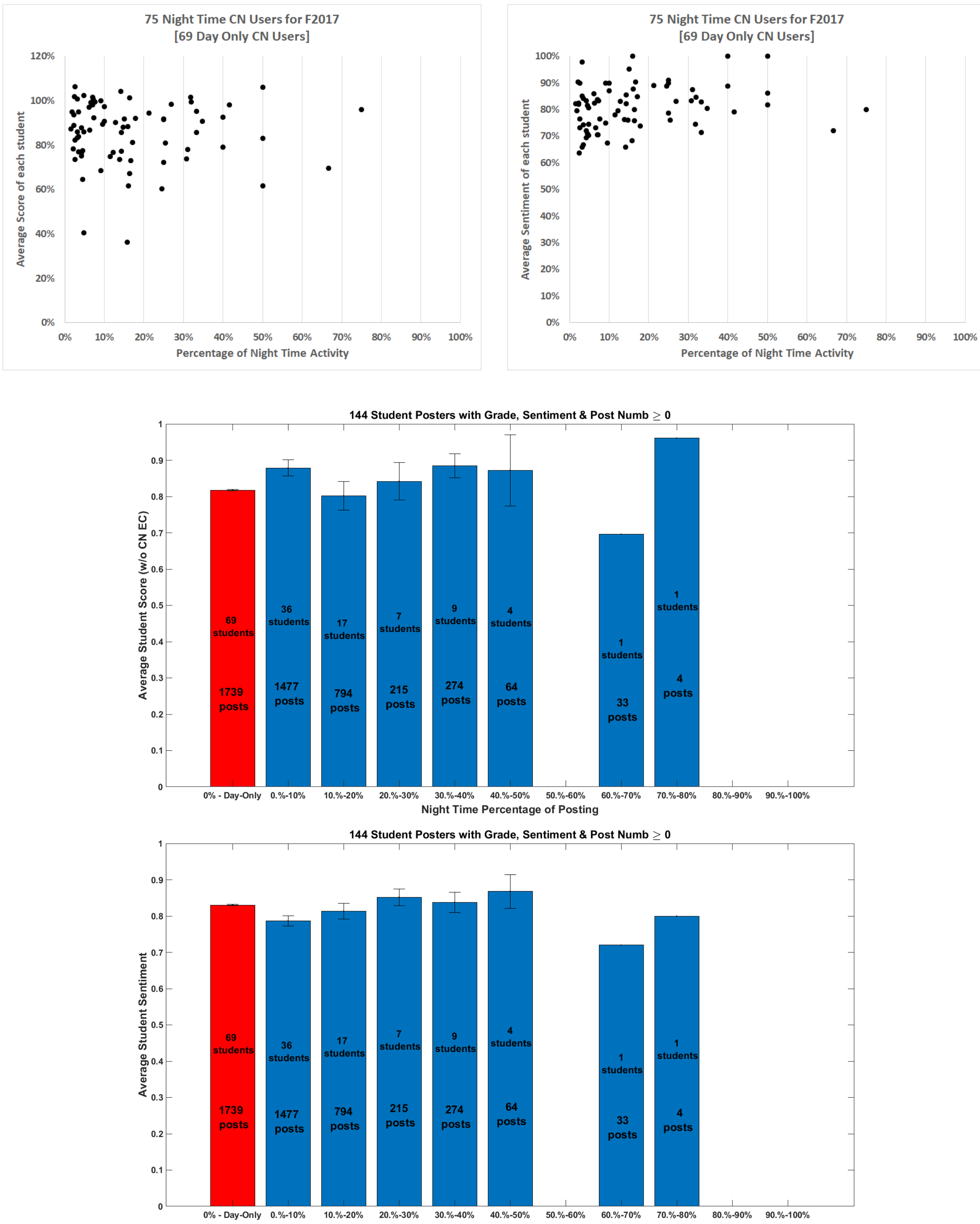


Figure 3.17: Average course score and sentiment scatter plot and bargraphs for each students versus their percentage of night time activity in Fall 2017

Chapter 4

Conclusion

4.1 Preface

The answers to the three questions in Chapter 3 reveal no statistically significant or important differences between the subgroups of CN users. We tested for gender differences, dissimilarities among CN activity by number of posts produced per student, and any variation in hourly time of student CN usage. All results failed to reject the null hypothesis, that there was any significant difference between our populations.

The aforementioned results demonstrate the subtlety when studying human behavior. There are many factors that influence student attitudes and course performance; each student is different and comes from a different background. Therefore, each semester is never truly identical to the previous one. However, the questions we posed are still valid. Had we discovered a consistent difference between the subgroups for any research question, this could indicate an underlying problem that certain types of students on CN experience. For example, suppose we had detected a noticeable difference in sentiment for the second most active quartile (Q2), less than all other quartiles of student activity, and time of day also had a discernible lower average sentiment for the latest night-time users. Suppose in this situation, the gender difference was indistinguishable. Upon further investigation into this dissimilarity, the instructor may have discovered that a small group of men, too small to observe in the gender study, works jobs during the evenings when most students study and do homework. This subgroup would have had to work at night to cover materials, in order to perform well in the course. They'd use CN primarily at night to vent their frustration and receive assistance from peers later in the day. The course scores wouldn't have reflected any differences yet CN provided a way to uncover students that otherwise would have gone unnoticed by the instructor.

The hypothetical example above illustrates how social media represents students. Social media has been found to reflect the students' personalities and views [33], and as such, CN recorded aspects of students' course views, opinions and impressions through student discussions. In other words, the CN data reflects the introductory physics course affect. Our results demonstrate, to the extent of R's Syuzhet sentiment analysis, that no disparity lies in the use of CN and ultimately implies that the course affect for males or females and heavy, light, or night-time leaning CN users were statistically equivalent between these groupings.

The outcomes of this study align with our initial goals for implementing CN into the introductory physics course. The idea was for it to enhance engagement, namely that students could share knowledge and interests, excitement and enthusiasm, worries and concerns, and build identity as individuals. As an instructor, it is important to probe these factors, the ease of which is now becoming more and more computationally accessible. The advent of massive open online courses (MOOCs) and the availability of online course tools such as Canvas, an open-source learning management system (LMS) used at IUPUI, provides assessments and analysis as to how often students' log-on, as well as their duration using the software. Twitter is another free tool that instructors have used that provides the capability to monitor limited 240 character comments called 'tweets' and engage students to communicate and share enthusiasms about their course. Many tutorials on text mining twitter feeds are easily found online. Stemming from those tutorials, the tools used in this work open an avenue for instructors to assess the affect of their class via CN. Measured affect that reveals overall positive and uniform sentiment between groups of students diminishes the need of the instructor to emphasize the affect aspect of the course and can free the instructor, with more confidence, to focus on providing course content.

To this point, this study has weighed the affect of the class by using CN, a media application whereby students can express themselves, and compare differences in student identities and activity. We discovered no such differences, even with a change from Fall 2014 course material and a different instructor in the Spring 2017 semester. This reveals equality in both sentiment and course score between gender, activity and night time usage. That is not to say that the current implementation of CN will be different in other classes or universities. But as for a subject with large gender differences [34] and taught by two different instructors, originating from different educational backgrounds and having different teaching practices, it was not exceptionally hard to arrange equality between these three CN data subgroups, as we have seen from our analysis.

4.2 Discussion

Data mining is a tool we used in the framework of physics education. Like any tool, it has its applicability and limits with regards to the accuracy of its measurements. People remain best at coding and categorizing qualitative data, such as our textual student posts. Computers, on the other hand, reign at performing repetitive tasks which is powerful when the data is large or in the extremes of big data, unmanageable by bare human analysis. The pitfalls of automated analytical processes lie in the lack of reasoning when fulfilling its functions. The oversight of a human is crucial when testing the reliability of the outputs and interpreting its conclusions, hence the adage "the computer is only as smart as the person using it". We will consider some drawbacks and difficulties in reaching our results and how these will influence the future direction of this study.

4.2.1 Word Frequency

Word frequency, obtained from running the code in Appendix B, revealed the most dominant words in our sentiment analysis. Unintentionally dominant words were first brought to light by using word clouds, an easy plot to produce in R. Word clouds graphically illustrate the most common words by placing them towards the center and in larger fonts, and colors for aesthetic purposes. When we first looked over the word clouds, words such as 'physics', 'problem', 'force' and 'sin' (deconstructed word from 'sin(x)') appeared front center in the word clouds. A quick look at the nrc word-sentiment lexicon revealed sentiment connected to these words. The associated sentiment to these words that we discovered didn't link to the actual meaning of the words in the context of this course. It was important to address this because these words had a definite impact on our sentiment analysis since students used these words frequently. We accounted for this inadvertent effect by adding these words to our masking words, words that are excluded from the sentiment analysis. Through a rather simple plotting tool that in itself doesn't reveal quantitative information, we were made aware of unintended effects on our quantitative analysis. The words clouds with proper masking now applied are viewable in the Appendix.

We have also neglected a part of common internet slang; many emotions are now conveyed with emojis or emoticons, namely characters that appear to simulate facial expression such as :-) or :) for happy, :'-(or :'(for crying, :-/ or :/ for skeptical/annoyed/undecided/uneasy/hesitant, ;-) or ;) for wink/smirk. Emotional context is also conveyed via computer lingo, aka non-standard words either pronounceable such as 'luv' or non-pronounceable such as 'lmao', along with hashtags and punctuations. All these were excluded in our sentiment analysis, since the sentiment lexicons do not contain any of these internet lingo. By finding a way to incorporate these, through the discovery or invention of self-made lexicon to be implemented alongside the 'nrc' lexicon, we can improve the precision on measuring affect of students.

4.2.2 Gender Study

Physics is largely a male dominated subject. To that effect, we wanted to observe if there is a difference between the larger male population and females. Many findings have found a difference in gender; differences have been observed between performance on standardized physics concept inventories [35] and also on students' course performance measured by their exam scores [36, 37, 38]. Our findings show no significant differences between gender. The uniqueness of our study is that we examined not only the course performance as measured by the students' overall course score, finding no difference between the two genders, but we peered further into the affect and found that the sentiment, as expressed on CN, shows no gender differences. It appears that CN is an equal playing field for both males and females. The course was monitored by the instructor during the course and loosely read through during analysis. We didn't find any exceptional gender distinctions on how students interact on CN nor certain words being predominantly used by one gender, as was studied by Bamman, Eisenstein and Schnoebelen [39] who investigated clusters of gender patterns and words in Twitter feeds. CN has consistently been verified in this study to be a fair arena for both women and men throughout the semesters. This shows that CN is a good platform to use for students, regardless of a predominant population of one gender.

4.2.3 Activity Study

Our study into the activity differences between students confirms the work done by Lau et al [40]. They found that the use of social media for academic purposes was not an indicator of academic performance. Similarly, we find no difference in our introductory physics courses among CN users by quartile.

The surprising fact is that those students who use CN more, share common sentiment and scores to those that use CN less. The implication either there is no extra benefit for students to use CN or it may boost those students who use CN less to an equal standing with those that more inclined to participate more heavily on this extracurricular platform. Ultimately, students that use CN to a larger degree aren't dominating the class sentiment and performing similarly to students that aren't using CN as extensively. The takeaway is that there was no discovered negative influence of CN in the use of these courses.

4.2.4 Time of Day Study

When one thinks of a student up at 3am in the morning, it's not a great leap to assume that this student is probably gaming, binge watching or doing last minute assignments or cramming for an exam. When we observed posts submitted in the late night/early time, which we specified to 11pm-6am, we were exploring the strong presumption that late night students would perform and probably feel differently to more consistent day-time students.

The results revealed that time, in fact, was not a strong indicator of either sentiment nor course score. It's important to interpret the results correctly by not placing too much emphasis on the few students that participated rarely, predominately at night. That was the motivation to create a threshold value of 6 posts, which we noticed as a maximum range for the first quartile of student activity in the previous section. The students that fell inside the first quartile with the lowest activity posted chiefly once with a peak of 6 posts. We noticed that these one-time or few posts were in response to the instructor encouraging students to introduce themselves at the start of the course. That is where a few of the really late night students appear in the results. We removed them since they do not play an essential role of addressing the underlying research question about consistent night-time users and their comparison with day-time users.

The incentive to investigate the effect of time on student sentiment and scores is again to see where CN is effective as a learning tool for students. Students again express similar sentiment, even if they are more night time users of CN. And their overall course performance showed no differences. This means CN can be used day or night, without loss of generality when averaged over the entire course.

4.3 Future Work

We plan to progress with topic modelling to extract automatic subject matter from the textual corpus.

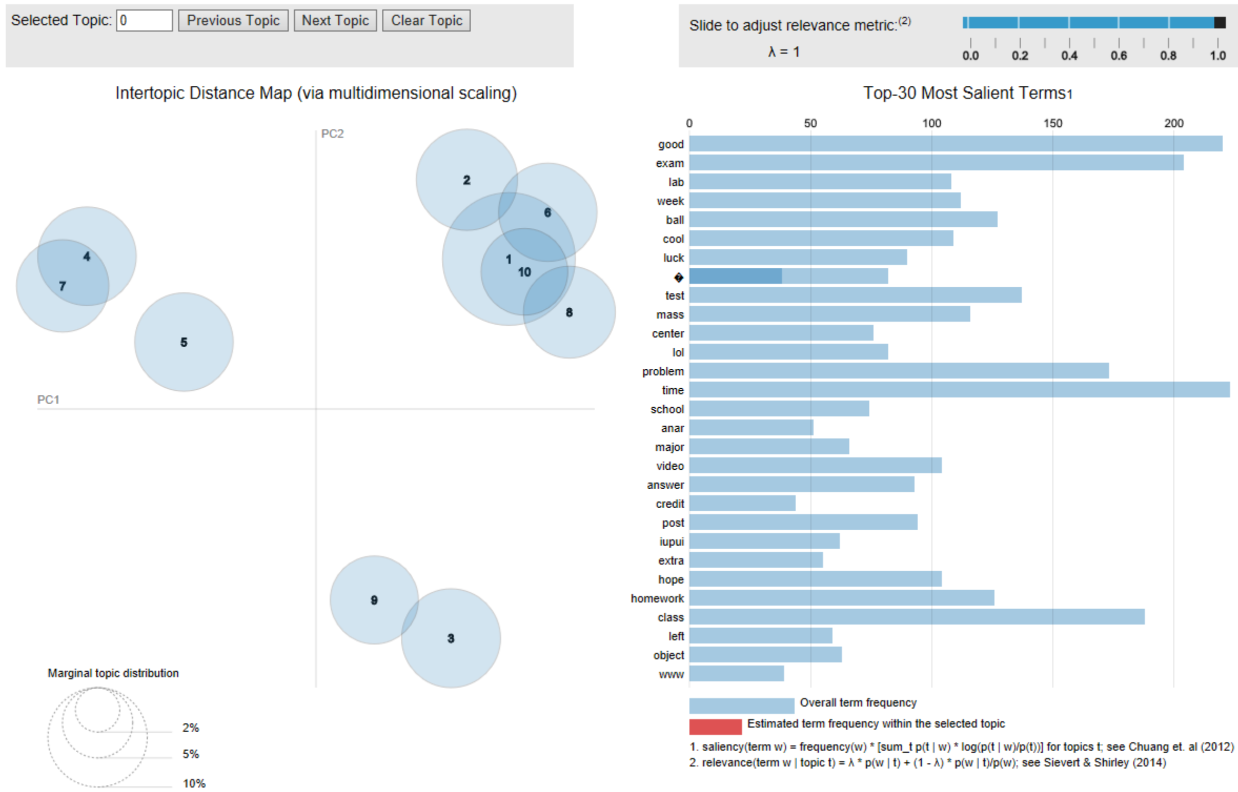


Figure 4.1: Preliminary LDAvis results for Fall 2015

Topic modelling (TM) is a statistical means to extract the “topic” or theme of the textual data. The primary method for TM counts frequency of the words and determines the probability of finding words as well as probabilistic relations with other words. TM can find semantics through wordnets. Wordnets are large lexicons that group sets of nouns, adjectives, verbs and adverbs together by synonyms and basic concepts. These word sets are called synsets. R can perform this with the package ‘wordnet’.

Another form of TM is LDA. LDA is one of the most popular tracks for generative statistical models. Fundamentally this model uses Bayesian inference, often applying the technique of Gibbs sampling. A common variation of Gibbs sampling, used by the R package ‘lda’, is collapsed Gibbs sampling. From this, R has the capability to feature the results by creating an interacting display by generating a JavaScript Object Notation (JSON) with changeable parameters, the number of topics and the weight of relevance for the topics denoted by λ that better controls the ranking of the topic relevance. Figure 4.1 illustrates the preliminary work using the package 'LDAvis' done on the word corpus from Fall 2015. With this we can numerically find out which words are used, extract meaning from these combination of words and quantitatively find out how often students use these words and in effect, how often certain topics of discussion appear in the semester. Examples of these topics could be group meetings to do homework or review for exams, give reassurances to peers before and after exams, provide interesting materials relevant to the course or simple dialogues about miscellaneous topics. We can view these topics as classifications and consider them from the perspective of gender, activity or time of day.

More in the distant future, we will attempt to link this with sentiment analysis and topic modelling and possibly also network analysis. There is ongoing research into the combination of sentiment analysis and topic modelling, such as Aspect-Sentiment Unification Model (ASUM) and Joint Sentiment/Topic Model (JSTM). We can pursue possible collaboration or incorporation of their techniques. This direction is more open ended. We hope to be able to piece these tools together with continuing experience with data mining and more methods applied to analyzing our textual data.

Inspired by the overall positive sentiment, we didn't dive into the negative aspects of the detected sentiment. The course was monitored frequently by the instructor but we can look at data more closely to see if there was any harassment or trolling. We can use the observed negative sentiment, since it consists around 20% of the post and is more manageable, to filter the comments and focus solely on any negative interactions that may have been propagated. Additionally, we have two more semesters to add. We can attempt to cluster our current and additional data with topic modelling to observe any distinction that men and women may have had from one another. We mentioned in gender studies discussion that a study found that choice of words are correlated to gender[39]. These genderized words impact the sentiment, since the sentiment is extracted from individual words. Future gender studies could use this to normalize the sentiment and increase sensitivity in observing gender differences. It has been reported in Gender Matters [34], that inequality and bias has been a driving force in discouraging women from pursuing physics. We will be vigilant in our continued study on gender.

The results from the student activity by quartiles provide evidence that a social media tool doesn't hinder the class performance in our context. However multi-tasking between academic tasks such as studying or paying attention in class and using electronic devices such as smartphones and laptops has been correlated with worse academic performance [40, 41, 42, 43, 44]. Whether CN is detrimental when students use it to multi-task is hard to measure but we could look at the times posts were recorded and coincide them with the times of lectures, recitations and labs. We can then compare the percentage of in-class CN usage to see if CN is harmful to student course sentiment or scores, like with nightly CN usage. If multi-tasking on CN confirms the prior studies, then this signals a possible intervention for our instructors to prohibit or limit mobile device usage in our introductory physics course.

Lastly, we will also experiment with other programming languages and their available packages and tools. Python provides many libraries for data mining tools. We want to compare the different applicable tools with those attainable with R for validity purposes. An analogous example would be the two different detectors, ATLAS and CMS, at the Large Hadron Collider (LHC) at CERN. They both have slightly different architectural designs, corresponding to python and R, but they measure the same phenomenons and are used to validate an observation. Additionally as we delve into further study of new possible tools for data mining, we may find packages only available from certain programming languages.

4.4 Conclusion

We obtained a null result for all our research questions. This may be viewed as unexciting or uneventful but the implications of this is positive in our class. It indicates that students based on gender, their CN activity or their temporal CN usage habits have no noticeable impact on their academic success. Our investigation made sure that all groups were feeling and performing equally; that we were not ignoring the impacts of the classroom on any minority group, whether that would be the lower number of females, lower number of CN usage, or lower number of nightly posts.

The significance isn't really the results but the novel methodology to examine a new avenue of student activity in the course. The use of CN allows data to be uninvasively collected in real-time, providing a unique way to tap into students' state of mind. Through sentiment analysis, we probe an innovatively different path to student affect since affect plays a significant role in a students' ability to learn. Analysis shows that it isn't adverse to certain groups of students and reveals an overall positive attitude. This proves that CN, as an educational social networking service, is not a disadvantageous course tool. It provides a common means for students to form connections to fellow classmates and express enthusiasm or apprehensions which can be alleviated by collective consensus.

This can be demonstrated by a student giving a CN post review on November 22, Fall 2015:

> "I like CN now, because it lets me know how my classmates have been doing, especially in the course...The close connection with others and updates are important especially in a large class; I, too wish my other classes will do the same...sometimes, I asked a question on concourse, but nobody responded to me."

CN provides a way for instructors to validate their assumptions on the course affect, and can thus use this in conjunction with what course content to cover or repeat when students express difficulties to their peers, when otherwise possibly shy or hesitant to confront the instructor. The application of data mining on students' intercommunication will continue to grow in importance as digital social networking advances into more and more areas of human interaction and likewise the tools to analyze natural language improve. We will expand on the available tools of data mining to examine our textual data and glean further and deeper insight into students' response to the implementation of a social networking resource on the course and their learning.

Bibliography

[1] Gregor M Novak, Evelyn T Patterson, Andrew D Gavrin, and Wolfgang Christian. Just in time teaching, 1999.

[2] Catherine H Crouch and Eric Mazur. Peer instruction: Ten years of experience and results. *American journal of physics*, 69(9):970–977, 2001.

[3] Edward F Redish, Jeffery M Saul, and Richard N Steinberg. Student expectations in introductory physics. *American Journal of Physics*, 66(3):212–224, 1998.

[4] Rosalind W Picard, Seymour Papert, Walter Bender, Bruce Blumberg, Cynthia Breazeal, David Cavallo, Tod Machover, Mitchel Resnick, Deb Roy, and Carol Strohecker. Affective learning—a manifesto. *BT technology journal*, 22(4):253–269, 2004.

[5] Beverly Woolf, Winslow Burleson, Ivon Arroyo, Toby Dragon, David Cooper, and Rosalind Picard. Affect-aware tutors: recognising and responding to student affect. *International Journal of Learning Technology*, 4(3-4):129–164, 2009.

[6] Anne Fernald. Intonation and communicative intent in mothers' speech to infants: Is the melody the message? *Child development*, pages 1497–1510, 1989.

[7] Alice M Isen and Barbara Means. The influence of positive affect on decision-making strategy. *Social cognition*, 2(1):18–31, 1983.

[8] Hyun Jung Oh, Elif Ozkaya, and Robert LaRose. How does online social networking enhance life satisfaction? the relationships among online supportive interaction, affect, perceived social support, sense of community, and life satisfaction. *Computers in Human Behavior*, 30:69–78, 2014.

[9] Winneke A van Der Schuur, Susanne E Baumgartner, Sindy R Sumter, and Patti M Valkenburg. The consequences of media multitasking for youth: A review. *Computers in Human Behavior*, 53:204–215, 2015.

[10] Shui-I Shih. A null relationship between media multitasking and well-being. *PLoS One*, 8(5):e64508, 2013.

[11] Mark W Becker, Reem Alzahabi, and Christopher J Hopwood. Media multitasking is associated with symptoms of depression and social anxiety. *Cyberpsychology, Behavior, and Social Networking*, 16(2):132–135, 2013.

[12] Kathryn R Wentzel. Student motivation in middle school: The role of perceived pedagogical caring. *Journal of educational psychology*, 89(3):411, 1997.

[13] Alan D Baddeley. The psychology of memory. *New York*, 1976.

[14] George Mandler. *Mind and body: Psychology of emotion and stress.* WW Norton & Company Incorporated, 1984.

[15] Barry Kort, Rob Reilly, and Rosalind W Picard. External representation of learning process and domain knowledge: Affective state as a determinate of its structure and function. In *Workshop on Artificial Intelligence in Education (AI-ED 2001), San Antonio,(May 2001)*, pages 64–69, 2001.

[16] Peter Ractham and Daniel Firpo. Using social networking technology to enhance learning in higher education: A case study using facebook. In *2011 44th Hawaii International Conference on System Sciences*, pages 1–10. IEEE, 2011.

[17] CourseNetworking LLC. Coursenetworking website, 2011.

[18] Adrienne Traxler, A Gavrin, and Rebecca Lindell. Networks identify productive forum discussions. *Physical Review Physics Education Research*, 14(2):020107, 2018.

[19] R Foundation for Statistical Computing. R programming language website, 1993.

[20] Matthew Jockers. Package 'syuzhet'. *URL: https://cran. r-project. org/web/packages/syuzhet*, 2017.

[21] Bing Liu, Minqing Hu, and Junsheng Cheng. Opinion observer: analyzing and comparing opinions on the web. In *Proceedings of the 14th international conference on World Wide Web*, pages 342–351. ACM, 2005.

[22] Minqing Hu and Bing Liu. Mining and summarizing customer reviews. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 168–177. ACM, 2004.

[23] Finn Årup Nielsen. A new anew: Evaluation of a word list for sentiment analysis in microblogs. *arXiv preprint arXiv:1103.2903*, 2011.

[24] Christopher Manning, Mihai Surdeanu, John Bauer, Jenny Finkel, Steven Bethard, and David McClosky. The stanford corenlp natural language processing toolkit. In *Proceedings of 52nd annual meeting of the association for computational linguistics: system demonstrations*, pages 55–60, 2014.

[25] Richard Socher, Alex Perelygin, Jean Wu, Jason Chuang, Christopher D Manning, Andrew Ng, and Christopher Potts. Recursive deep models for semantic compositionality over a sentiment treebank. In *Proceedings of the 2013 conference on empirical methods in natural language processing*, pages 1631–1642, 2013.

[26] Saif M Mohammad and Peter D Turney. Emotions evoked by common words and phrases: Using mechanical turk to create an emotion lexicon. In *Proceedings of the NAACL HLT 2010 workshop on computational approaches to analysis and generation of emotion in text*, pages 26–34. Association for Computational Linguistics, 2010.

[27] Andrew Ortony and Terence J Turner. What's basic about basic emotions? *Psychological review*, 97(3):315, 1990.

[28] R Plutchik. A general psychoevolutionary theory of emotion in r. plutchik & h. kellerman (eds.) emotion: Theory, research, and experience (vol. 1, pp. 189-217), 1980.

[29] Ann Bowling. Mode of questionnaire administration can have serious effects on data quality. *Journal of public health*, 27(3):281–291, 2005.

[30] Adrienne L Traxler, Ximena C Cid, Jennifer Blue, and Ramón Barthelemy. Enriching gender in physics education research: A binary past and a complex future. *Physical Review Physics Education Research*, 12(2):020114, 2016.

[31] A. Gavrin and Rebecca Lindell. Students' course based social media participation is positively correlated with homework effort. *AAPT Phys. Ed. Res. Conf*, 2016. Sacramento, CA.

[32] John W Tukey. Comparing individual means in the analysis of variance. *Biometrics*, pages 99–114, 1949.

[33] Tomas Chamorro-Premuzic. How different are your online and offline personalities. *The Guardian*, 2015.

[34] Jennifer Blue, Adrienne L Traxler, and Ximena C Cid. Gender matters. *Physics Today*, 71(3):40–46, 2018.

[35] James Day, Jared B Stang, NG Holmes, Dhaneesh Kumar, and DA Bonn. Gender gaps and gendered action in a first-year physics laboratory. *Physical Review Physics Education Research*, 12(2):020104, 2016.

[36] Eric Brewe, Vashti Sawtelle, Laird H Kramer, George E O'Brien, Idaykis Rodriguez, and Priscilla Pamelá. Toward equity through participation in modeling instruction in introductory university physics. *Physical Review Special Topics-Physics Education Research*, 6(1):010106, 2010.

[37] Steven J Pollock, Noah D Finkelstein, and Lauren E Kost. Reducing the gender gap in the physics classroom: How sufficient is interactive engagement? *Physical Review Special Topics-Physics Education Research*, 3(1):010107, 2007.

[38] Akira Miyake, Lauren E Kost-Smith, Noah D Finkelstein, Steven J Pollock, Geoffrey L Cohen, and Tiffany A Ito. Reducing the gender achievement gap in college science: A classroom study of values affirmation. *Science*, 330(6008):1234–1237, 2010.

[39] David Bamman, Jacob Eisenstein, and Tyler Schnoebelen. Gender identity and lexical variation in social media. *Journal of Sociolinguistics*, 18(2):135–160, 2014.

[40] Wilfred WF Lau. Effects of social media usage and social media multitasking on the academic performance of university students. *Computers in human behavior*, 68:286–291, 2017.

[41] Aryn C Karpinski, Paul A Kirschner, Ipek Ozer, Jennifer A Mellott, and Pius Ochwo. An exploration of social networking site use, multitasking, and academic performance among united states and european university students. *Computers in Human Behavior*, 29(3):1182–1192, 2013.

[42] R Arteaga Sánchez, Virginia Cortijo, and Uzma Javed. Students' perceptions of facebook for academic purposes. *Computers & Education*, 70:138–149, 2014.

[43] Minas Michikyan, Kaveri Subrahmanyam, and Jessica Dennis. Facebook use and academic performance among college students: A mixed-methods study with a multi-ethnic sample. *Computers in Human Behavior*, 45:265–272, 2015.

[44] Reynol Junco. Student class standing, facebook use, and academic performance. *Journal of Applied Developmental Psychology*, 36:18–29, 2015.

Appendix A

Code: Sentiment Analysis

```
library(syuzhet)
library("tm")
library("SnowballC")
require(xlsx)

##################################################
############# STEP 1: IMPORTING ##################
##################################################
year = 2015
yearstr <- as.character(year)
studentdata <- read.xlsx("data/studentcndata.xlsx", sheetName = yearstr)

messages <- studentdata$text

##################################################
############# STEP 2: PREPROCESSING ##############
##################################################

################ Text word stemming/cleaning ################
comments.eng <- messages

comments.eng <- tolower(comments.eng) #to lowercase
comments.eng <- gsub("\\d","",comments.eng) #deletes numbers
comments.eng <- gsub("","",comments.eng) #deletes apostr
comments.eng <- gsub("[[:punct:]]","␣",comments.eng) #delete punctuation
comments.eng <- gsub("[[:cntrl:]]","␣",comments.eng) #delete white space
comments.eng <- gsub("^[[:space:]]+","",comments.eng) #delete begin. space
comments.eng <- gsub("[[:space:]]+$", "",comments.eng) #delete end space

doc.list <- strsplit(as.character(comments.eng),"[[:space:]]+")

##################################################
############# STEP 3: STOP WORDS #################
##################################################

################ SetUp Stopwords ################
stpwords <- read.csv("data/stopwords.txt", header=FALSE)
maskwords <- read.csv("data/maskwords.txt", header=FALSE)
colnames(stpwords) <- c("words")
colnames(maskwords) <- c("words")
#### Append all words together ####
stop_words <- NULL
stop_words <- c(as.character(stpwords$words),as.character(maskwords$words),stopwords("
    english"),stopwords("SMART"))

#### Create stemmed and/or stopped word eliminated list -> message ####
message <- NULL
for (var in 1:length(doc.list)) {
```

```
  ### Apply Stop Words with stemming ###
  # doc.list.stemmed <- wordStem(unlist(doc.list[[var]]),language = "english")
  # doc.list.stemmed <- tm_map(Corpus(VectorSource(doc.list.stemmed)), removeWords, as.
      character(unlist(stop_words)))
  # message[[var]] <- matrix(doc.list.stemmed)

  ### Apply Stop Words without stemming ###
  doc.list.unstemmed <- tm_map(Corpus(VectorSource(unlist(doc.list[[var]]))), removeWords,
      as.character(unlist(stop_words)))
  message[[var]] <- matrix(doc.list.unstemmed)

}


##################################################
############ STEP 4: SA ########################
##################################################

######### INITIALIZE ALL VARIABLES ###############

WORDCORPUS <- list() #filtered words from ind. posts
wordpost_list <- list() #word-lexicon matches
sentiment_list <- list() #position of sent. word match
emotion_list <- list() #position of emot. word match

positivewordhit <- list() #pos. words in post
negativewordhit <- list() #neg. words in post

wordhitcount <- 0 #numb. of word-lexicon matches


##### WORD-LEXICON MATCHES FOR EACH POST #########
for(i in 1:length(message)) {

  #### filter words from indiv. post #####
  x = message[[i]][1]
  if(!identical(x[[1]], character(0))){
    for(j in 1:length(x[[1]])){
      if(x[[1]][j]=="" && !is.na(x[[1]][j])){x[[1]][j]<-NA}
      }
    w <- na.omit(x[[1]])
    attributes(w)$na.action <- NULL
  } else {
    w <- NULL
  }

  ##### Sentiment & Emotion Analysis #####
  if(length(w)!=0){
    sentiment_w <- get_sentiment(w, method="nrc")
    emotion_w <- get_nrc_sentiment(w);
    wordpost_w <- w[as.logical(sentiment_w)]
  } else {
    sentiment_w <- ""
    positivewordhit[[i]] <- ""
    negativewordhit[[i]] <- ""
    wordpost_w <- ""
  }

  ###### Add to List #####
  WORDCORPUS[[i]] <- w
  wordpost_list[[i]] <- wordpost_w
  sentiment_list[[i]] <- sentiment_w
  emotion_list[[i]] <- emotion_w


  ### POSITIVE & NEGATIVE SENT. WORD LIST ###
  poshit <- ""
```

```
  neghit <- ""
  for(j in 1:length(sentiment_w)){
    if(sentiment_w[j] == 1){
      poshit <- c(poshit,w[j])
      wordhitcount <- wordhitcount+1
    }else if(sentiment_w[j] == -1){
      neghit <- c(neghit,w[j])
      wordhitcount <- wordhitcount+1
    }
  }
  # Number of Pos & Neg Words:
  positivewordhit[[i]] <- poshit
  negativewordhit[[i]] <- neghit

}

### All Unlisted Pos. and Neg. Words from data #####
poswordhits <- unlist(positivewordhit, recursive = TRUE)
negwordhits <- unlist(negativewordhit, recursive = TRUE)


################ COUNT SENTIMENT ##################
sentiment_matrix <- NULL
for(s in 1:length(sentiment_list)){
  if(length(sentiment_list[[s]])==1 && sentiment_list[[s]]==""){
    sentiment_matrix[s] <- 0
  } else{
    sentiment_matrix[s] <- sum(sentiment_list[[s]])
  }
}
sentiment_matrix <- as.matrix(sentiment_matrix)
colnames(sentiment_matrix) <- "totalsent" #column name

################ COUNT EMOTION ###################
emotion_matrix <- NULL
for(j in 1:length(emotion_list)){
  emot <- as.matrix(emotion_list[[j]])
  emot <- colSums(emot)
  emotion_matrix <- rbind(emotion_matrix,emot)
}
emotion_matrix <- as.matrix(emotion_matrix)
rownames(emotion_matrix) <- c() #remove row names

##################################################
############# STEP 5: SAVING SA RESULTS ###########
##################################################

### APPEND Sentiment TO studentdata ####
studentdata <- cbind(studentdata, sentiment_matrix)

### APPEND Emotion TO studentdata ####
studentdata <- cbind(studentdata, emotion_matrix)

##################################################

#### SAVE DATA ####
write.xlsx(studentdata,"results/studentcndataresults.xlsx", append=TRUE, sheetName = yearstr
    , row.names=TRUE)
```

Appendix B

Word Clouds

```
library("wordcloud")
library("RColorBrewer")
######################################################################
####################### POSITIVE ###################################
######################################################################
posCorpus <- ""
posCorpus <- Corpus(VectorSource(poswordhits))

dtm <- TermDocumentMatrix(posCorpus)
m <- as.matrix(dtm)
v <- sort(rowSums(m),decreasing=TRUE)
d <- data.frame(word = names(v),freq=v)
head(d, 10)
png(filename=paste(paste("results/",paste("positivewordcloud",yearstr,sep = ""),sep = ""),".png",sep = ""))
set.seed(1234)
wordcloud(words = d$word, freq = d$freq, min.freq = 1,
          max.words=100, random.order=FALSE, rot.per=0.35,
          colors=brewer.pal(8, "Dark2"))
dev.off()
write.csv(d, file = paste(paste("results/poswordhitfreq",yearstr,sep=""),".csv",sep=""))
######################################################################
####################### NEGATIVE ###################################
######################################################################
negCorpus <- ""
negCorpus <- Corpus(VectorSource(negwordhits))

dtm <- TermDocumentMatrix(negCorpus)
m <- as.matrix(dtm)
v <- sort(rowSums(m),decreasing=TRUE)
d <- data.frame(word = names(v),freq=v)
head(d, 10)
png(filename=paste(paste("results/",paste("negativewordcloud",yearstr,sep = ""),sep = ""),".png",sep = ""))
set.seed(1234)
wordcloud(words = d$word, freq = d$freq, min.freq = 1,
          max.words=100, random.order=FALSE, rot.per=0.35,
          colors=brewer.pal(8, "Dark2"))
dev.off()
write.csv(d, file = paste(paste("results/negwordhitfreq",yearstr,sep=""),".csv",sep=""))
######################################################################
######### ALL FILTERED POSTED WORDS (WORDCORPUS) ##################
######################################################################
wordCorpus <- ""
wordCorpus <- Corpus(VectorSource(WORDCORPUS))

dtm <- TermDocumentMatrix(wordCorpus)
m <- as.matrix(dtm)
v <- sort(rowSums(m),decreasing=TRUE)
d <- data.frame(word = names(v),freq=v)
head(d, 10)
png(filename=paste(paste("results/",paste("allfilteredwordcloud",yearstr,sep = ""),sep = ""),".png",sep = ""))
wordcloud(words = d$word, freq = d$freq, min.freq = 1,
          max.words=100, random.order=FALSE, rot.per=0.35,
          colors=brewer.pal(8, "Dark2"))
dev.off()
write.csv(d, file = paste(paste("results/allwordhitfreq",yearstr,sep=""),".csv",sep=""))
```

Table B.1: Fall 2014 and Fall 2015 Word Clouds

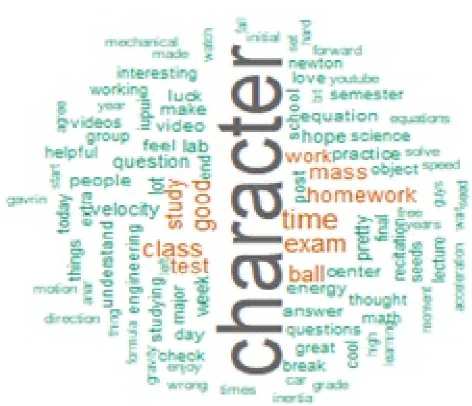
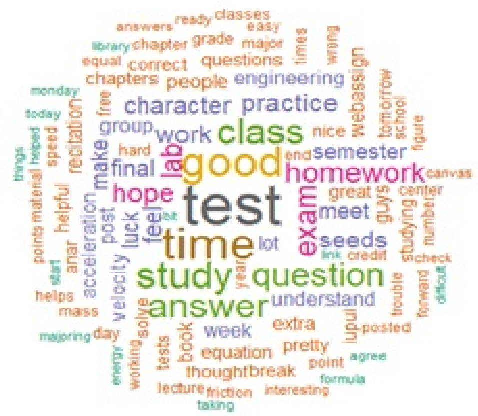
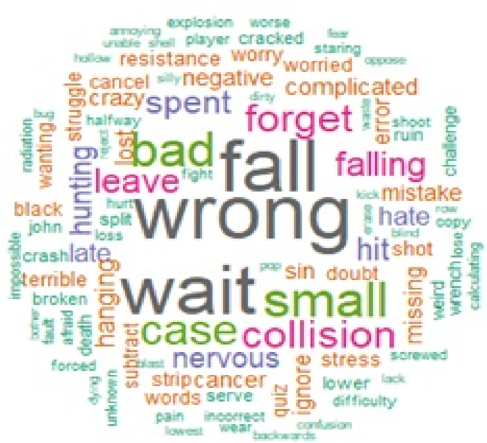
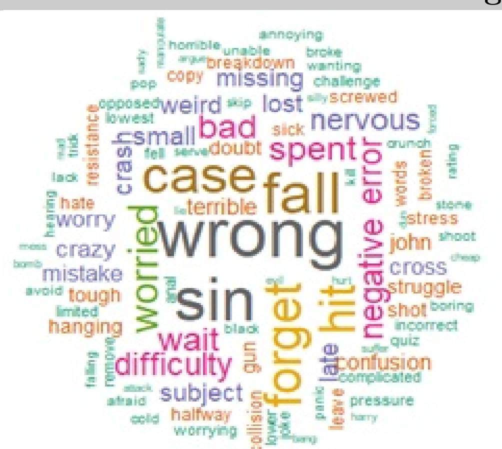
Fall 2014	Fall 2015
Most Common Words	
	
Most Positive Words	
	
Most Negative Words	
	

Table B.2: Fall 2016 and Spring 2017 Word Clouds

<table>
<tr><th>Fall 2016</th><th>Spring 2017</th></tr>
<tr><th colspan="2">Most Common Words</th></tr>
<tr><td>
</td><td>
</td></tr>
<tr><th colspan="2">Most Positive Words</th></tr>
<tr><td>
</td><td>
</td></tr>
<tr><th colspan="2">Most Negative Words</th></tr>
<tr><td>
</td><td>
</td></tr>
</table>

Appendix C

Stop Words

```
ll
ill
didn
www
don
a
about
above
across
after
afterwards
again
against
all
almost
alone
along
already
also
although
always
am
among
amongst
amoungst
amount
an
and
another
any
anyhow
anyone
anything
anyway
anywhere
are
around
as
at
back
be
became
because
become
becomes
becoming
been
before
beforehand
behind
being
below
beside
besides
between
beyond
bill
both
bottom
but
by
call
can
cannot
cant
co
computer
con
could
couldnt
cry
de
describe
detail
do
done
down
due
during
each
eg
eight
either
eleven
else
elsewhere
empty
enough
etc
even
ever
every
everyone
everything
everywhere
except
few
fifteen
fify
fill
find
fire
first
five
for
former
formerly
forty
found
four
from
front
full
further
get
give
go
had
has
hasnt
have
he
hence
her
here
hereafter
hereby
herein
hereupon
hers
herse"
him
himse"
his
how
however
hundred
i
ie
if
in
inc
indeed
interest
into
is
it
its
itse"
keep
last
latter
latterly
least
less
ltd
made
many
may
me
meanwhile
might
mill
mine
more
moreover
most
mostly
move
much
must
my
myse"
name
namely
neither
never
nevertheless
next
nine
no
nobody
none
noone
nor
not
nothing
now
nowhere
of
off
often
on
once
one
only
onto
or
other
others
otherwise
our
ours
ourselves
out
over
own
part
per
perhaps
please
put
rather
re
same
see
seem
seemed
seeming
seems
serious
several
she
should
show
side
since
sincere
six
sixty
so
some
somehow
someone
something
sometime
sometimes
somewhere
still
such
system
take
ten
than
that
the
their
them
themselves
then
thence
there
thereafter
thereby
therefore
therein
thereupon
these
they
thick
thin
third
this
those
though
three
through
throughout
thru
thus
to
together
too
top
toward
towards
twelve
twenty
two
un
under
until
up
upon
us
very
via
was
we
well
were
what
whatever
when
whence
whenever
where
whereafter
whereas
whereby
wherein
whereupon
wherever
whether
which
while
whither
who
whoever
whole
whom
whose
why
will
with
within
without
would
yet
you
your
yours
yourself
yourselves
```

Appendix D

Masking Words

sin
physics
problem
problems
force
units
meters
seconds
kilograms
order
of
magnitude
vectors
components
kinematics
particle
position
displacement
speed
velocity
average
instantaneous
derivative
integral
acceleration
free fall
relative
reference
frame
projectile
trajectory
uniform
circular
motion
tangential
centripetal
acceleration
angular
period
dynamics
mass
newton
laws
newton
newtons
principle
superposition
net
weight
tension
normal
spring
constant
hooke
hookes
law
free-body
free
body
diagram
diagrams
kinetic
static
coefficient
friction
terminal
speed
centripetal
center
mass
work
joule
electron-volt
electron
volt
work-kinetic
energy
theorem
scalar
product
power
watt
potential
conservative
non-conservative
nonconservative
total
mechanical
energy
stable
unstable
linear
momentum
conservation
elastic
inelastic
collisions
collision
impulse
impulse-momentum
center-of-mass
center
of
mass
frame
rocket
equation
angular
displacement
velocity
acceleration
moment
inertia
rotational
kinetic
energy
parallel
axis
theorem
torque
moment
arm
rolling
without
slipping
vector
product
angular
momentum
conservation
angular
momentum
orbit
kepler
keplers
universal
gravitation
escape
velocity
gravitational
field
statics
static
equilibrium
conditions
stress
strain
youngs
young
modulus
pressure
hydrostatic
pascal
archimedes
archimedess
principle
buoyant
incompressible
volume
flow
rate
continuity
bernoulli
equation
simple
harmonic
motion
restoring
amplitude
frequency
hertz
angular
phase
pendulum
small
angle
approximation
harmonic
waves
longitudinal
transverse
wavelength
wave
number
linear
mass
density
interference
constructive
destructive
standing
waves
node
antinode
fundamental
harmonics