What Determines Involvement at Carnegie Mellon?

Ellie Gurary, Bruce Jackson, Maggie Soderholm, Chrissy Swierkocki, Jen Sung

Section 1: Introduction

1.1 Research Question and Motivation

Carnegie Mellon is not known for its school spirit. In particular, we have noticed low attendance at sporting events and other traditional campus activities. The creation of the Tartan Rewards Program and the Loyal Scott Program, which aim to increase attendance at Carnegie Mellon events, illustrate the lack of student initiative to be present at such activities. Understanding the level of involvement of various groups on campus can be used to unify the student body and improve attendance at school events. Carnegie Mellon sports teams, clubs, and incentive programs, such as Tartan Rewards, could be possible clients of this survey because the information could improve their attendance rates. This study needs to be done now to benefit incoming classes of students and attract more students to come to Carnegie Mellon.

1.2 Brief Literature Review

Significant research has not been conducted in this area of interest, particularly about Carnegie Mellon. We did, however, come across a few studies and reports that bear a resemblance to what we are interested in. A few items from Carnegie Mellon, including the Tartan and an independent study, looked at the relationship between social life and major/area of study as well as how the mascot has improved attendance at sporting events and other on-campus activities. We also explored a few external sources. "A Journey through Adult Student Involvement on Campus" looks at the affects of involvement on development and learning. It states "the greater the student's involvement in college, the greater will be the amount of student learning and personal growth" (p. 307)." (See links below).

1.3 Quick Summary of Main Results

We prepared and distributed a paper and pencil survey and administered this survey to classes sampled from the Spring 2011 Carnegie Mellon Undergraduate Class Schedule. We asked questions about students' time commitment in on and off campus organizations and predictor variables to identify biases. From the results of our study, we found that Age, Campus Housing, Campus Job, Club Sports, Paid Other Position (not listed in the survey), and Class Years have significant effects on the Involvement Score. We also compared other independent variables to our overall Involvement Score. In the end, although we were not able to develop a robust model, we were able to describe involvement on Carnegie Mellon's campus and how this information can be of value to increase school spirit and Tartan pride.

Section 2: Methods

2.1 Target Population and Frame

Our target population is Carnegie Mellon University undergraduate students. Given that our mode of data collection is a random cluster sample of undergraduate university wide courses, our sampling frame is all undergraduate students in these randomly selected courses (which are our clusters).

2.2 Sample Size

The population (N) is 5,705. We are using a Margin of Error (ME) of .05 and a worst-case standard deviation (SD) of .5. From this we calculated n_0= 385. Because we are conducting SRS without replacement, we found n≥360.6= 361. And because we are doing clustered sampling, we inflated this figure by 20% to obtain n=433.2=433. To compensate for non-response, we divided our sample size by the individual non-response rate, which we found to be approximately 85% due to the low in-class non-response rate and the high attendance reported in each class. This gave us our maximum sample size of 433/.85 = 510. We assumed that the average class size was 25 students (CMU Admissions). To find how many clusters we needed in order to reach our calculated sample size of 434, and the number of professors we need to contact, we divided our sample size by the average class size, 510/25 = 20.4. We assumed a professor non-response rate of approximately 50%, so we divided the amount of clusters by .5 to get 41 professors we needed to contact.

2.3 Sample Design and Methods

We did a random cluster sample of undergraduate classes at Carnegie Mellon University to survey our target population of undergraduate students. We used a list of the undergraduate level classes offered for the Spring 2011 semester, and using a random number generator, we selected 41 classes from the list. We used the random number generator to choose a number, and we then went through the list, which was organized alphabetically so the academic departments were intermixed, and counted to the number and selected the class that was landed on. We then emailed the professors individually asking to come to their class for a few minutes to administer a brief survey. After receiving permission from the professor, we went to 20 classes and surveyed the students in attendance. When we administered the survey, we gave a brief introduction to the content of the survey and our motivation for the survey.

We decided on SRS without replacement because we would like to give equal chances to all the undergraduate classes. Since our target population is Carnegie Mellon undergraduates, it would be best to not stratify in order to give all the undergraduates an equal chance. We chose to do SRS without replacement since once we choose a class for our sample we are not replacing the chosen class back into our random number generator.

this doesn't make sense

We contacted the randomly chosen professors by email (See Appendix A). We sent individual emails to professors and sent the first batch of 26 emails out. We emailed the professors in batches at the end of the week in order to survey the classes the week after. If a professor did not respond within a week, we sent a reminder email to account for the nonresponse. After 2 reminder emails we planned to meet the professors face-to-face, but by the time we got to the list of nonresponse professors, we already reached and surpassed our sample size of 434.

We administered in-class surveys of randomly selected classes from the undergraduate course offerings. We felt that this was the best method of collecting data because it will minimize coverage errors and lead to the best random sample of undergraduate students with the highest response rate. Since the survey is handed out in the class and typically the students do not leave until dismissed by the professor, we anticipated a very high response rate from the students who were in class that particular day. We did not ask any class identifiers or keep the surveys organized by class in order to preserve anonymity.

Our survey consists of the informed consent statement on the first page, and then our questionnaire (See Appendix B). The questionnaire consisted of three categories, including demographic questions, predictor variables, and questions that lead to our creation of an "Involvement Score", which will be discussed later. The demographic questions ask the students' gender, year, college within Carnegie Mellon and race. The predictor variables ask about specific activities or affiliations that a student can be a part of on and off campus. Some examples of these questions are "Are you a Resident Assistant?" and "Do you have a campus job?" The predictor variables also consist of quantitative questions, such as asking how much time they spend on academics outside of classes and how many sports events they have attended. The involvement score questions ask students how many organizations they are a part of and how many hours they spend on those organizations (See Appendix C).

a brief discussion of how to compute the involvement score would be good here.

2.4 Response

Though we predicted a response rate of only 50% from professors, we emailed 41 and heard back from 32 for a 78.05% response rate. Of those who responded, 7 declined to grant us their class time, for a refusal rate of 17.07%. Only 21.95% of the emailed professors did not respond at all.

Within each class we saw a high response rate as well. Classes typically had very high attendance, according to the professors, and almost all students in attendance completed the survey. We approximate an 85% response rate due to the high rates of attendance and survey completion.

Demographics wise, our sample was fairly representative. We had 59.38% males, which is near Carnegie Mellon's total population average of 57% males. We were able to sample students from every college as well. The break down was as follows: CIT 31.32%, H&SS 19.68%, MCS 14.46%, SCS 14.46%, CFA 11.24%, Tepper 7.43. When we compared this break

down to the actual distribution of students across colleges, we were pleased that the sample was representative. We also were able to get students from all years of study. We had 21.20% Freshman, 33.06% Sophomores, 18.46% Juniors, 26.17% Seniors, and 1.22% Fifth years. This distribution is not as representative of the actual distribution, which shows more equality across classes. One reason we believe we had so many Sophomores and Seniors was due to specific classes we sampled. One class was only for CIT Sophomores and the Beatles class is difficult to get into unless you have an early scheduling time, such as rising Seniors. See Figure 2.

2.5 Post-Survey Processing

Each completed survey is entered into a spreadsheet so we can use our variables to produce a per person and aggregate involvement score that measures participation in campus events and activities. To account for nonreponse items, we categorized them as MCAR (Missing Completely At Random), so we just did a case-wide deletion. When we tried to impute the nonresponse items, it was difficult because hot deck imputation was impossible since so many respondents had similar results, while regression imputation was also as difficult because of the many outliers we had (thus inflating our mean).

there should be info about nonpartic. in non-resp.

As mentioned before, we did not keep our clusters separate after we gathered our data (in order to preserve anonymity). In hindsight, since we did a random cluster sample, we should have kept the clusters separate in order for us to calculate the cluster mean and cluster variance to keep in line with our cluster sampling structure. Therefore, our data analysis was done on individuals (which may add errors into our analysis).

Section 3: Results

3.1 General Results

We found that many of the questions that were asked in the survey resulted in very similar answers over most of the students in our sample. Either Carnegie Mellon undergraduate students are not particularly involved, or we were asking questions targeted at a subsection that we either missed in surveying, or more likely is simply smaller than we had anticipated. We attempted multiple statistical analyses, including various models, removal and inclusion of strongly significant individual respondents. We found during this analysis that the variation between models in terms of strength was relatively small. The adjusted R-Squared value between models varied only a few percentage points between the different types of models we attempted. The final model that resulted from our surveying efforts was not very strong, and did not explain much of the variability in the Involvement Score variable.

greater pre-testing could help with this

3.2 Statistical Analyses

We ran our regression in R. First, we calculated the Involvement Score by compiling all

the responses from the first couple of questions. We summed the values dealing with the number of organizations that each respondent was a member of, as well as what positions they held and how involved they were in multiple aspects. This was then treated as a response variable and multiple regressions were run. Though multiple types of models were attempted, the multivariate linear regression model below was deemed the best. It incorporates the parameters which were significant in the full model, and the adjusted R^2 of the models did not change much between models, it varied by a few percentage points at most. We found that the R^2 was not very high for any of the models, indicating that the parameters we had data for were not accounting for much of the involvement score variable. In fact, our final model predictor variables account for 15.63% of the total variation in involvement score.

Our regression model:

what can you say about this model? were all coef's significant? can you interpret the coefs? etc.

Involvement Score = 131.3986 – 1.5580 (Age) -25.898(Not in Campus Housing) – 26.0857 (In Campus Housing) – 22.2024 (No Campus Job) – 16.4310 (Campus Job) + 2.5663 (Sophomore) + 9.1545 (Junior) + 7.7254 (Senior) + 2.6495 (5th Year Senior) – 36.9486 (No Club Sports) – 35.0398 (Club Sports) + 8.4523 (No Paid Other Position) + 27.0887 (Paid Other Position) – 17.7156 (Not in Special Interest Housing) – 13.5849 (Special Interest Housing)

(See Appendix D)

3.3 Conclusions about our Research Questions

Our model accounts for an adjusted 15.63% of the variability in Involvement score. This model is not very strong at predicting the Involvement Score based on the questions that we asked. The research question of what contributes to undergraduate involvement at Carnegie Mellon has not adequately been explored at this point in our research. A less quantitative approach might result in a stronger outcome. Different questions must be explored to better answer the research question, and perhaps a different format ought to be sought. The multiple choice and yes/no survey that was put together seems to have either been too structured, or was asking the wrong questions, to get at the heart of the issue.

Section 4: Discussion

4.1 Our Research Questions

Our research question asks what are the factors that motivate CMU undergraduates' involvement on campus? We looked into factors such as how many organizations a student is a member of, and how active they are in each of these organizations. These factored into an Involvement Score, and we used demographic variables such as school, and response variables such as how many majors the student has, if they have specific campus involvement, such as

working as a teacher's assistant, or as an orientation counselor. We looked to see which factors were statistically significant in affecting student involvement, how significant they were, and how they affected involvement.

4.2 Surprising/Unexpected Results

shoudl be specific about what biases you think are there, and why.

One aspect of our results that surprised us was the low R^2 value. There were many biases that we could not take into account with the questions we asked on the survey. Another surprising result was that the Greek or non-Greek variable was not significant. The data show that Greeks are still very involved on campus, but with the inherent biases in the data, we are not able to determine if it impacts involvement. We were also surprised that the number of social networking sites may be positively correlated with Involvement Score, since the amount of time spent on-line would detract from time spent working with organizations or campus life. This would make sense, though, if these people used the social networking sites as recruitment tools for their organizations.

4.3 Brief Answers to Research Questions

Our model accounts for an adjusted 13.98% of the variability in Involvement score. This model is not very strong at predicting the involvement score based on the questions that we asked. The research question of what contributes to undergraduate involvement at Carnegie Mellon has not adequately been explored at this point in our research.

4.4 Strengths

Our data collection had many successful aspects. We were able to collect over 450 surveys, and participants expressed minimal confusion about the questions. Our data was representative of the CMU population on all demographic aspects, with the possible exception of year. We also had a much higher response rate than expected, with the professor response rate 78% and the student response rate approximately 85%.

4.5 Weaknesses

We asked questions that were either too structured in students' ability to answer in a meaningful way, or were not asking about the correct topic. We had some early miscommunication with professors about gathering data, which could have potentially led to an increase in bias of the results. Also, we had some surveys that were filled out with clearly illegitimate answers. Some students were in more than one randomly selected class, and though some made this known to the surveyors, potentially there were subjects taking the survey more

what kind?

than once. Also, the number of students in the randomly selected classes greatly varied, from as little as three students to over one hundred and fifty students.

the main difficulty here would be SE's that are too small

Because we did not keep the clusters after we gathered the data, our analysis could be skewed towards individuals. Also, we were not aware of how to do regression models on clusters, so we decided to keep them as individuals.

Another weakness in our study is our model itself. We have a fairly low R^2, which is expected since we are running a social science study. But, we did not account in out regression model for endogeneity bias, or any instrumental variables we could implement from our data. Thus, we have a sparsely set of variables that are significant and others that we thought should be (like Greek, as mentioned previously) that aren't.

4.6 Take Home Message

Overall, we feel that the questions and methods of our research were not the most effective in explaining the variability we saw in the Involvement Score. There is room for improvement in continuing efforts to explore undergraduate involvement at Carnegie Mellon, and it is important to continue researching this important issue.

List of References

Carnegie Mellon University. *Break Down of Social Life at Carnegie Mellon* [Data File]. Retrieved from http://www.studentsreview.com/specific_detail.php3?uid=1150&f=Social&d_short=CM&d_school=Carnegie+Mellon+University.

Chun-Mei Zhao, George D. Kuh and Robert M. Carini. (Mar. - Apr., 2005). A Comparison of International Student and American Student Engagement in Effective Educational Practices. *The Journal of Higher Education*. Vol. 76(2). Retrieved from http://www.jstor.org/stable/3838723?&Search=yes&searchText=university&searchText=student&searchText=engagement&list=hide&searchUri=/action/doBasicSearch%3Facc%3Don%26Query%3Dstudent%2Buniversity%2Bengagement%26gw%3Djtx%26acc%3Don%26prq%3Dcampus%2Bstudent%2Binvolvement%2Buniversity%26Search%3DSearch%26hp%3D25%26wc%3Don%26acc%3Don&prevSearch=&item=3&ttl=18975&returnArticleService=showFullText

Kenyon, Elizabeth L. (Jun., 1997). Seasonal Sub-Communities: The Impact of Student Households on Residential Communities. *The British Journal of Sociology* Vol. 48(2). Retrieved from http://www.jstor.org/stable/591753?&Search=yes&searchText=involvement&searchText=university&searchText=student&searchText=campus&list=hide&searchUri=/action/doBasicSearch%3Facc%3Don%26Query%3Dcampus%2Bstudent%2Binvolvement%2Buniversity%26gw%3Djtx%26acc%3Don%26prq%3Dcampus%2Bstudent%2Binvolvement%2Bengagement%26Search%3DSearch%26hp%3D25%26wc%3Don%26acc%3Don&prevSearch=&item=1&ttl=8004&returnArticleService=showFullText

Morgan, Wendy. A Journey Through Adult Student. Retrieved from http://www.colostate.edu/Depts/SAHE/JOURNAL2/2001/Journey.htm

Porter, Sabrina. (September 15, 2008). Scotty brings out spirit around campus. *The Tartan*. Retrieved from http://thetartan.org/2008/9/15/news/

Appendix A: Contact Letter to Professors

Professor,

My name is Jennifer Sung and I am a Junior Economics and Decision Science double major at Carnegie Mellon. Currently I am taking a course called Sampling, Survey and Society (36-303) with Professor Brian Junker. In this course we have a group project where we choose a topic/problem to work on throughout the semester.

Our group project is about the measurement of involvement on Carnegie Mellon's campus through on-campus organizations, leadership positions, athletics, etc. We are planning to survey the students in class, and we have chosen ~25 classes, all chosen randomly by a random number generator. Your class, [Enter course title], was chosen to be a part of our sample. We would like permission to come at the end of class to hand out the surveys to your students (it should take about 10 minutes at most). It is a pencil and paper administered survey, so we would be providing the materials ourselves.

We would greatly appreciate your cooperation so that we can move forward with our project (and hopefully get good samples). We were wondering if we could survey your class sometime this week or next week.

Again, we would really appreciate the opportunity to survey your class for our class project.

Sincerely,
Jen Sung

Appendix B: Informed Consent Statement

Study Title: Campus Involvement Survey

Purpose of the Study
This study is for our semester project in a statistics elective called Sampling, Survey, and Society with Professor Brian Junker. Our semester project entails us to administer surveys to collect data for our project. The purpose of our project is to measure student's involvement on campus.

Procedure
You will be asked to answer questions on the survey (next page).

Risks & Benefits
The risks and discomfort associated with participation in this study are no greater than those ordinarily encountered in daily life or during the performance of routine physical or psychological examinations or test. There are no personal benefits from your participation in this study.

Confidentiality
We will not be collecting any identifiers; therefore, you will be taking the survey anonymously.

Rights
This is all voluntary. You are free to not finish the survey at any point. This will not result in any penalty or loss of benefits or rights to which you might otherwise be entitled.

If you have any questions, please direct them to the administrators for your class or email the 36-303 instructor, Brian Junker (brian@stat.cmu.edu) .

By beginning the survey, you agree that above information has been explained to you and all your current questions have been answered. By starting the survey, you agree to participate in this study. Please turn the page to continue.

Appendix C: Questionnaire

CAMPUS INVOLVEMENT SURVEY

Please answer the following questions to the best of your ability according to the definitions below:

Organization: Voluntary, extra-curricular groups you are not compensated for involvement on campus (compensation includes credit, payment, housing and any other forms of compensation)

Time Period: Please answer all questions in reference to this past summer (Summer 2010), last semester (Fall 2010), this semester (Spring 2011), and this coming summer (Summer 2011).

How many organizations are you an active member of?: _____

How many organizations do/did you serve on the executive board for? _____

How many organizations do/did you hold a non-executive position in? _____

About how many hours a week do/did you spend on your organizations (including scheduled meetings)? _____

How many hours during the above time period have you spend recruiting for your organizations? _____

Do you currently have a social networking account? If so, on which website(s)?_____

How many Facebook friends do you have?
A: N/A
B: 1-200
C: 201-400
D: 401-600
E: 601-800
F: 801-1000
G: 1001-1200
H: 1201+

How many times a week do you tweet? (write N/A if you do not have a Twitter Account)
_____

How many athletic (varsity, IM and/or club) events have you attended as a spectator since the beginning of the school year?
A: 0
B: 1-3
C: 4-6
D: 7-9
E: 10-12
F: 13+

How many social events have you attended over the last month which are unrelated to academics?
A: 0
B: 1-4
C: 5-8
D: 9-12
E: 13-16
F: 17+

How did you meet your closest Carnegie Mellon friends? (Choose One)
A: Dorm
B: Classes
C: Non-Campus Organizations
D: Parties
E: Campus Events
F: Campus Organizations
G: Other _____

Are you Greek?
Y N
Are you a varsity athlete?
Y N
Are you involved in club sports?
Y N
Are you involved in IM sports that are not related to another organization you are a part of?
Y N
Are you a Teaching Assistant?
Y N

Are you a Resident Assistant?
Y N
Are you a compensated (i.e. credit, $) Research Assistant?
Y N
Are you a Community Advisor?
Y N
Are you a Mudge Mentor?
Y N
Are you a CMU Peer Tutor?
Y N
Are you an HOC?
Y N
If you are an HOC, how long have you had this position? _____ semester(s)
Are you a pre-college counselor?
Y N
Are you a pre-college assistant director?
Y N
Are you involved in Special Interest Housing (Including wellness housing, quiet housing, etc)?
Y N
Do you have a campus job (not listed above)?
Y N
Do you have an off-campus job?
Y N
Are you involved in off-campus organizations (ex: religious organizations, volunteer organizations, etc.)?
Y N
Are you involved in a paid or compensated position that is not listed above?
Y N
If so, please note:
_____
_____

On average, how many hours a week do you spend on academics? (Outside of class)
A: 0
B: 1-15
C: 16-30
D: 31-45
E: 46-60
F: 61-75
G: 76+

How old are you? _____

What class year are you?_____

How many majors do you have? _____

How many minors do you have (if any)? _____

Which of the 6 CMU Colleges are you in? (If you have more than one major, go by your primary major)
H&SS
SCS
CIT
TSB
CFA
MCS

Gender:
M F

What is your ethnicity/race?
A: American Indian or Alaskan Native
B: Asian (non-Pacific Islander)
C: African-American
D: Other Pacific Islander
E: White
F: Hispanic
G: Other_____

Are you currently living in campus housing?
Y N

Appendix D: Regression Model Statistics

Coefficients:	Estimate
Constant	131.399 (22.776)**
Age	-1.5580 (0.600)**
CampusHousingno	-25.898 (8.973)**
CampusHousingyes	-26.0857 (8.908)**
CampusJobno	-22.2024 (8.416)**
CampusJobyes	-16.4316 (8.502)
ClassYear2	2.5663 (2.463)
ClassYear3	9.1545 (2.978)**
ClassYear4	7.7254 (3.183)**
ClassYear5	2.6495 (8.782)
ClubSportsno	-36.9486 (13.297)**
ClubSportsyes	-35.0398 (13.498)
PaidOtherPostionno	8.4523 (7.583)
PaidOtherPostionyes	27.0887 (8.426)**
SIHno	-17.7156 (10.552)
SIHyes	-13.5849 (11.001)

Note: Numbers in Parentheses are standard errors. The (**) denote significant result. ClassYear is code for each class year: so ClassYear2 is sophomores, ClassYear3 is juniors, ClassYear4 is seniors and ClassYear5 is fifth years. SIH is Special Interest Housing.

Holding all other variables constant at zero:

- If a person is a freshman, they have an average expected involvement score of 131.3986
- If a person does not live in campus housing, they have an average expected involvement score of 105.5006

- If a person does live in campus housing, they have an average expected involvement score of 105.3129
- If a person does not have a campus job, they have an average expected involvement score of 109.1962
- If a person does have a campus job, they have an average expected involvement score of 114.9676
- If a person is a sophomore, they have an average expected involvement score of 133.9649
- If a person is a junior, they have an average expected involvement score of 140.5531.
- If a person is a senior, they have an average expected involvement score of 139.124.
- If a person is a 5th-year senior, they have an average expected involvement score of 134.0481.
- If a person is not involved with club sports, they have an average expected involvement score of 94.45.
- If a person is involved with club sports, they have an average expected involvement score of 96.3588
- If a person does not have a paid position not specifically asked about, their average expected involvement score is 139.8509.
- If a person has a paid position not specifically asked about, their average expected involvement score is 158.4873.
- If a person is not in special interest housing, their average expected involvement score is 113.683.
- If a person is in special interest housing, their average expected involvement score is 117.8137.

ANOVA

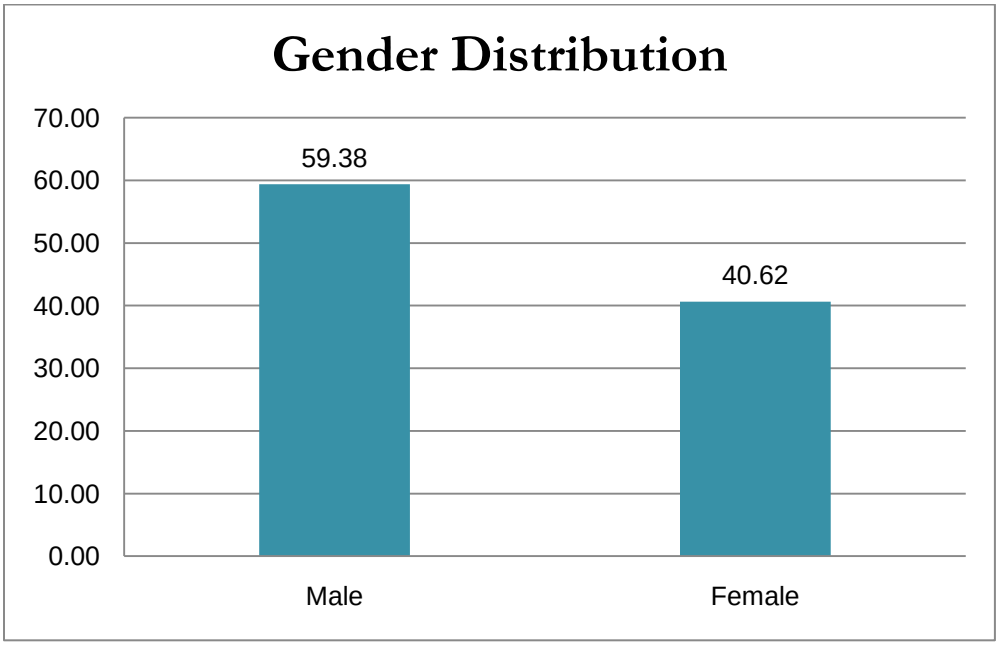
	Constant	Age	Campus Housing	Campus Job	Class Year	Club Sports	Paid Other Position	SIH
Sum of Squares	1.11	3556.17	8614	3758.83		2211.31	7126.53	1318.77
Deg. of Freedom	1	2		2	4	2	2	

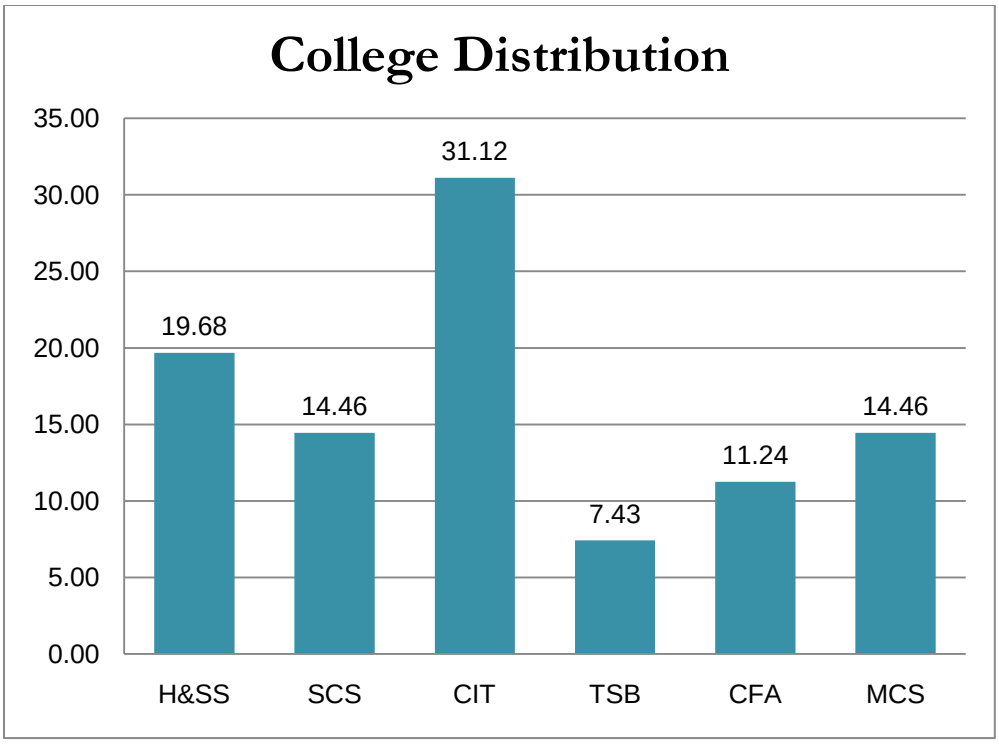
Residual standard error: 17.58235

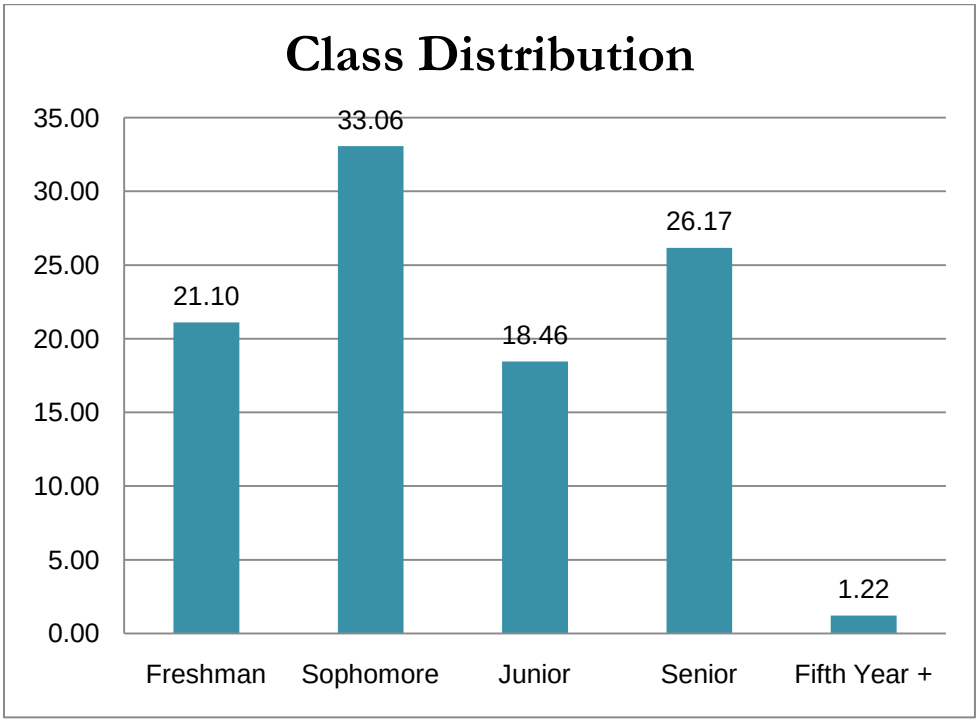
Figure 1: Continuous variables

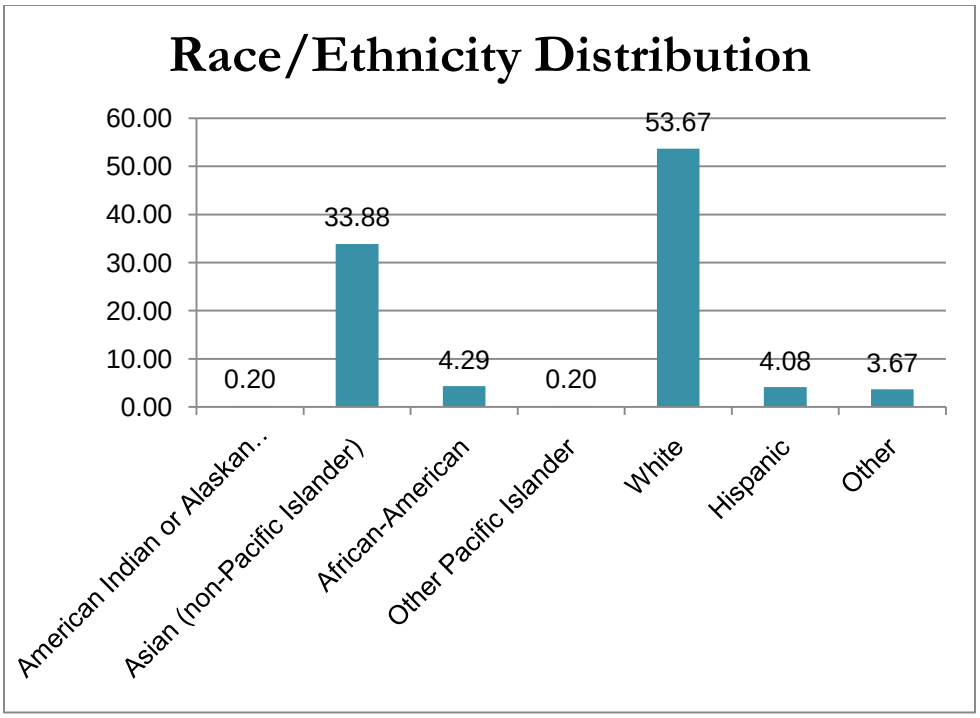
Variable	Minimum	Maximum	Median	Mean
Involvement	0	263	9	13.74
Number of Social Network Sites	0	5	1.29	1
Age	17	47	20	20.24
Class Year	1	5	2	2.56
Majors	0	3	1	1.3

Figure 2: Demographics (in percents)




College Distribution
35.00
30.00
25.00
20.00
15.00
10.00
5.00
0.00
19.68
14.46
31.12
7.43
11.24
14.46
H&SS
SCS
CIT
TSB
CFA
MCS


Class Distribution
35.00
30.00
25.00
20.00
15.00
10.00
5.00
0.00
21.10
33.06
18.46
26.17
1.22
Freshman
Sophomore
Junior
Senior
Fifth Year +



On Campus Housing	Count
Yes	289
No	154
N/A	6

Figure 3: Distribution of Involvement score

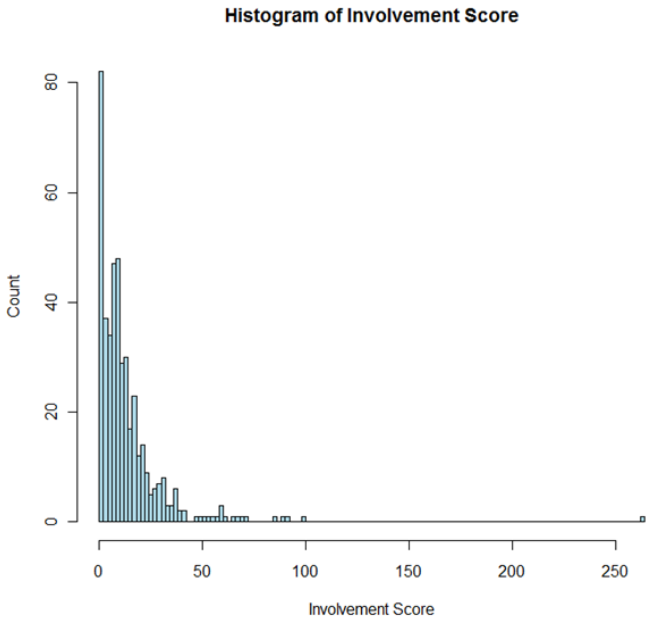


Figure 4: After removing the possible outlier, the distribution of Involvement score distribution looks like this:

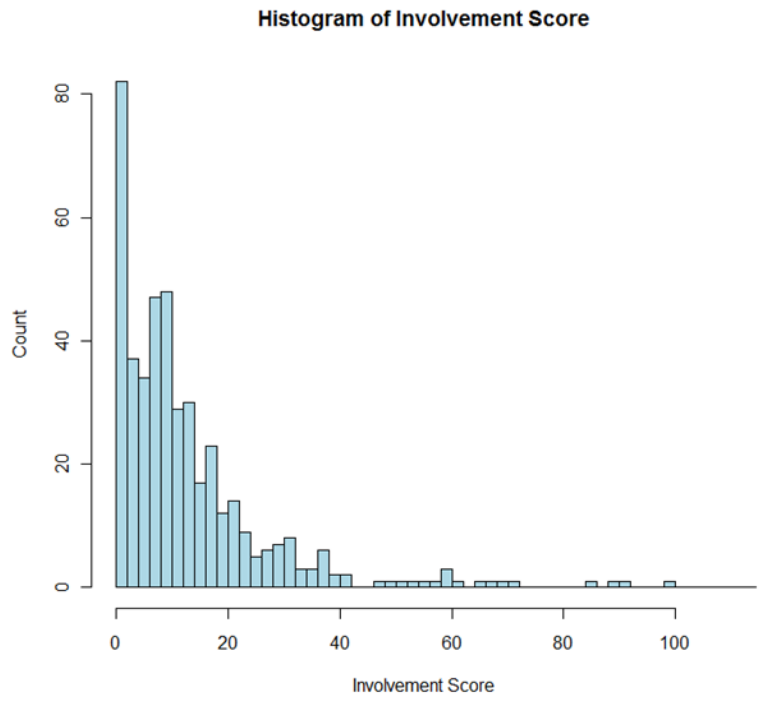


Figure 5: Relationship between Involvement score and the number of Athletic events attended

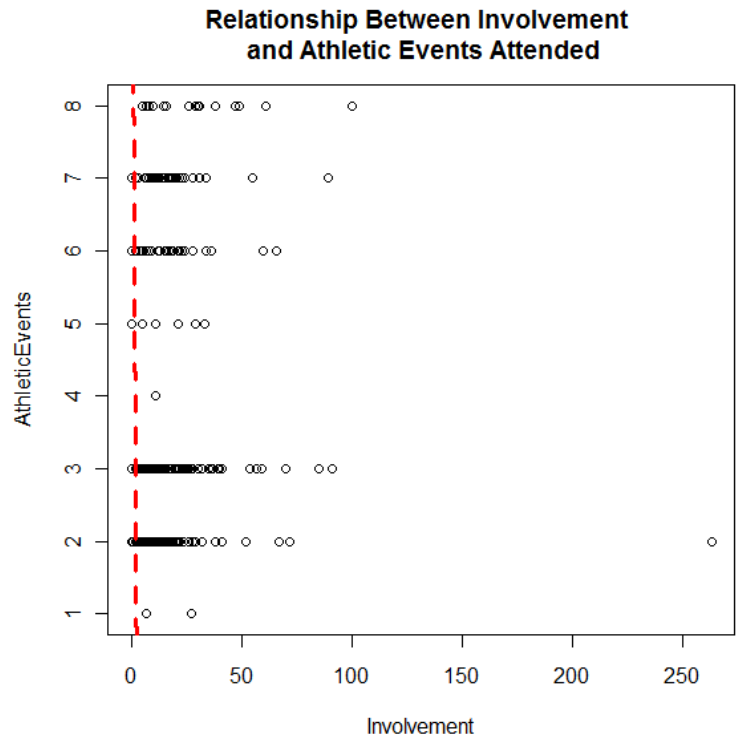


Figure 6: Fitted Values from Regression

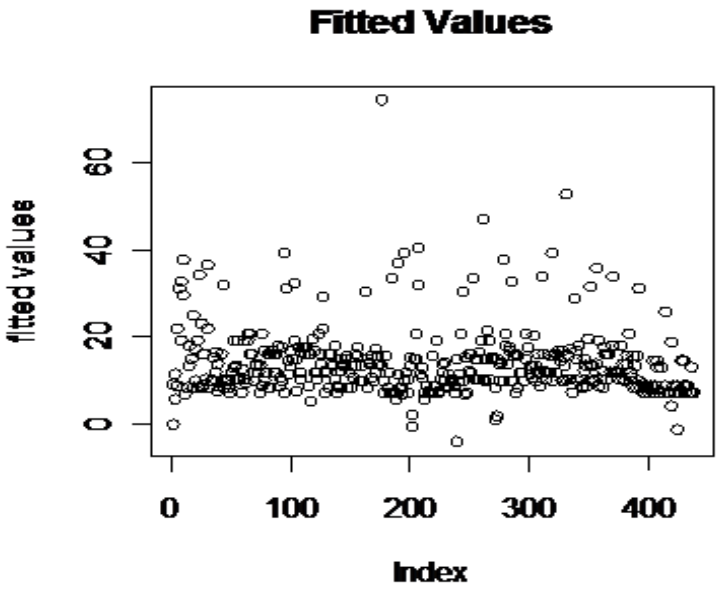
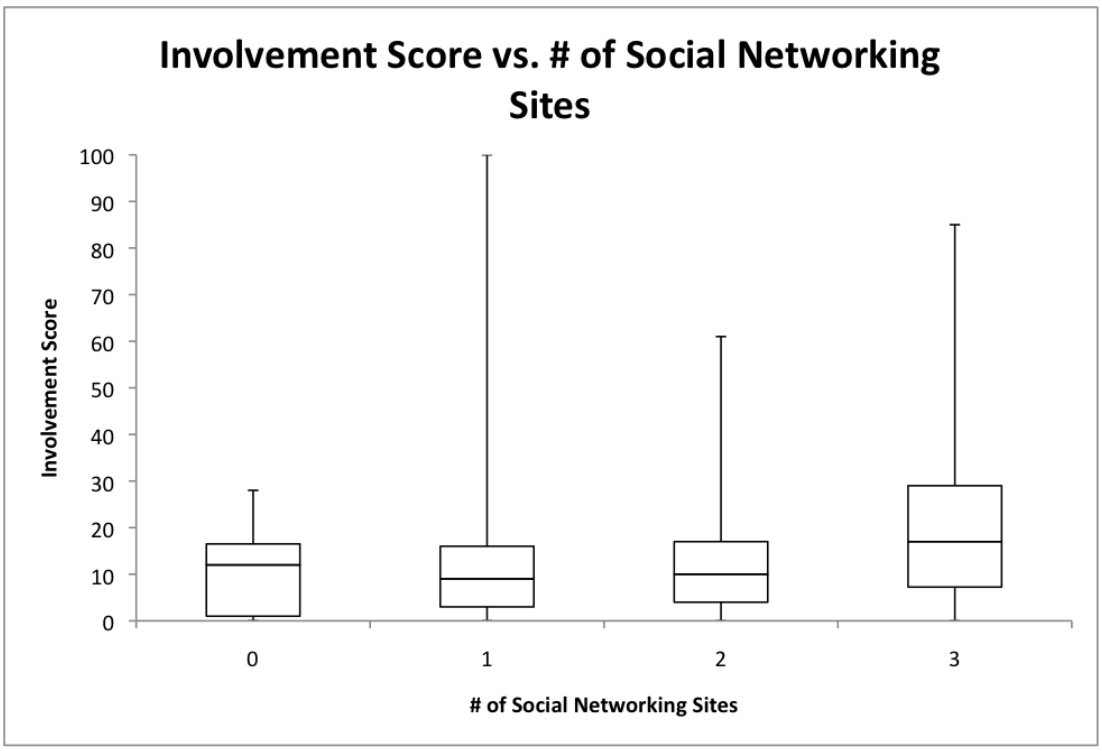


Figure 7: Boxplot of Involvement Score vs. Number of Social Networking Sites



R-Code Appendix:

```
rm(list = ls())
data <- read.csv("FinalData.csv")
summary(data)
attach(data)
names(data)

model1 <- lm(Involvement~ AcademicTime + Age+AthleticEvents+CA+
CampusHousing+CampusJob+ ClassYear+ +ClubSports+
CMUPT+College+Fbookfriends+Gender+ Greek+ HOC+IMSports+Majors+Minors+ MM+
Off.CampOrgs+Off.CampusJob+ PaidOtherPostion+Pre.CollegeAD+Pre.CollegeC+ RA+
Race+ ResearchA+ SIH+SocialEvents+ SocialNet+ TA+ tweet+ VarsityAthlete)
summary(model1)
aov(model1)

data<-read.csv("Workingdata.csv", header=T)
attach(data)

hist(Involvement, col="light blue", main="Histogram of Involvement Score",
xlab="Involvement Score", ylab="Count",breaks=100)
```

```
hist(Involvement, col="light blue", main="Histogram of Involvement Score",
xlab="Involvement Score", ylab="Count",breaks=100, xlim=c(0, 110))

line1<-lm(Involvement~AthleticEvents)
plot(Involvement, AthleticEvents, main="Relationship Between Involvement \n and Athletic
Events Attended")
abline(line1, col="red", lwd=3, lty="dashed")

line3<-lm(Involvement~ClassYear)
plot(Involvement, ClassYear, main="Relationship Between Involvement \n and Class Year")
abline(line3, col="red", lwd=3, lty="dashed")

line4<-lm(Involvement~Majors)
plot(Involvement, Majors, main="Relationship Between Involvement \n and Number of
Majors")
abline(line4, col="red", lwd=3, lty="dashed")


model1.logit <- glm(Involvement~ AcademicTime + Age+AthleticEvents+CA+
CampusHousing+CampusJob+ ClassYear+ +ClubSports+
CMUPT+College+Fbookfriends+Gender+ Greek+ HOC+IMSports+Majors+Minors+ MM+
Off.CampOrgs+Off.CampusJob+ PaidOtherPostion+Pre.CollegeAD+Pre.CollegeC+ RA+
Race+ ResearchA+ SIH+SocialEvents+ SocialNet+ TA+ tweet+ VarsityAthlete)
summary(model1.logit)

model2 <- lm(Involvement~ AcademicTime + Age+ CampusHousing+CampusJob+
ClassYear+ +ClubSports+ CMUPT+College+Fbookfriends+Gender+ Greek+
IMSports+Majors+Minors+ Off.CampOrgs+Off.CampusJob+ PaidOtherPostion+ RA+ Race+
ResearchA+ SIH+SocialEvents+ SocialNet+ TA+ tweet+ VarsityAthlete)
summary(model2)

model3 <- lm(Involvement~Age+CampusHousing+CampusJob+as.factor(ClassYear) +
ClubSports + PaidOtherPostion + SIH)
summary(model3)
aov(model3)
```