Bassem Mikhael
Christopher Loncke
Abigail Daughtrey
Prerna Agarwal
Yong-Gyun Choi
April 29, 2011

Final Report

Introduction:

Numerous health studies have demonstrated the damaging effects of excessive caffeine consumption on cardiac wellbeing in addition to psychological and mental health (Mayo, 2011). Significant concern has been expressed that students today consume large amounts of caffeine to keep up with their academic workload or to enhance their performance in sports activities (Erst, 2011). This problem is especially severe in schools which are more work-intensive, which Carnegie Mellon University certainly is. Therefore, we have studied patterns of caffeine consumption among undergraduate students at Carnegie Mellon with the hope of characterizing caffeine consumption, as well as gaining an understanding of students' perception of caffeine consumption. Through the former component, we hoped to determine if there are any correlations between the amount/type of caffeine consumption with GPA, the college the student is in, their busyness level, etc. On the other hand, the latter component is designed to test whether students are comfortable with the amount of caffeine they consume, and to understand the more subjective perspective on this matter.

Overall, we found that surprisingly, there is no correlation between caffeine consumption and a student's perception of whether caffeine was bad for them. We also found that students who believe that caffeine increases their performance tend to consume more caffeine. Further, we found that increased caffeine consumption is correlated with less sleep and more symptoms of caffeine addiction. The regression model shows that increasing the busyness score is also directly correlated with caffeine consumption.

Methods:

The target population for our study is comprised of all current Carnegie Mellon undergraduate students on the Pittsburgh campus. The sampling frame we selected to administer the study is C-book, a published booklet listing all current students' names, years, and email addresses. A very small number of students actively choose to opt-out of being included in the C-book; these students are members of the target population but not the sampling frame, and therefore will not be included in the sample, resulting in coverage error. Due to the very small number of students in this category, however, we believe the sampling frame to be an excellent tool to use to sample the population.

The design of the questions proposed for the study is essential, as the results will allow for an understanding of the sample participants and a determination of the caffeine consumption patterns at Carnegie Mellon University. For this survey, we decided to keep the questions short, simple and concise to streamline the process of responding for survey participants. Also, we decided to make all of the questions multiple choice to maintain consistency between answers and to avoid any discrepancies

which would arise from coding variable responses. The survey questions, which are listed in Appendix A, comprise five major categories: demographic information, academic performance, extracurricular involvement, general well-being, and student perception. The demographic questions asked for basic information such as class year, gender, college, etc. The academic performance questions determined the student's GPA, units taken, hours spent on class work, etc. This can allow us to correlate caffeine consumption with academic performance. Extracurricular involvement, the third category, asks questions which aim to determine how busy the student is; questions ask the student how many hours he/she spends on leisure, extracurricular activities, employment, etc. This data was translated into a "busyness" score and can be tested to determine correlation with caffeine consumption (see section 3 for a precise definition of this score). The fourth category of questions takes a slightly different approach, and aims to determine the student's general wellbeing through measures such as frequency of exercise and sleeping patterns. Lastly, the perception questions are designed to try to understand the students' subjective feelings and comfort level with their present level of caffeine consumption.

We found that the easiest and most efficient manner to administer this survey is to administer it online. In order to do this, we chose Google Forms, a free service which is part of GoogleDocs, rather than the more traditional choice of SurveyMonkey. Google Forms offered automatic data entry into a spreadsheet at no cost, while SurveyMonkey lacked these features. Both to minimize expense and to minimize the probability of human error which would occur from manually copying data into a spreadsheet, we chose to use Google Forms. After composing the survey questions, we pretested the survey with a group of students and discovered, in addition to minor errors, that we had to encode questions in a specific manner to allow for multiple responses. Certain questions in the survey necessitate the capability to choose multiple responses, so we modified the survey to allow for this.

The actual sampling method chosen was simple random sampling with stratification. First, the strata were decided to be the 5 possible class years, freshman, sophomore, junior, senior, and 5th year students, because we believed that this would allow us to obtain a smaller variance and that the classes would differ enough to make stratification effective. In order to calculate sample size for each stratum we used the mean error formula for stratified samples:

$$ME = (1.96) \times \sqrt{\sum_{h-1}^{H} W_h^2 \,(1 - f_h)\frac{s_h^2}{(f \times N_h)}}$$

We used 0.1 for the ME to give us a small confidence interval, and 2 for s based off of pretesting and the responses of our own group to the question "How many servings of caffeine do you consume on a daily basis?" W_h and N_h were both easily calculable using admissions statistics from the Carnegie Mellon website. Using the class populations on page 8 of the admissions statistics provided from CMU (CMU Admissions, 2011), we calculated f to be about 0.015. Using those same statistics, the respondents needed for each category were calculated to be 20 freshmen, 18 sophomores, 18 juniors, 17 seniors, and 3 fifth years.

Year	Number Needed	Number Selected
Freshmen	20	120
Sophomores	18	108
Juniors	18	108
Seniors	17	102
5th Years	3	18
Total	**76**	**456**

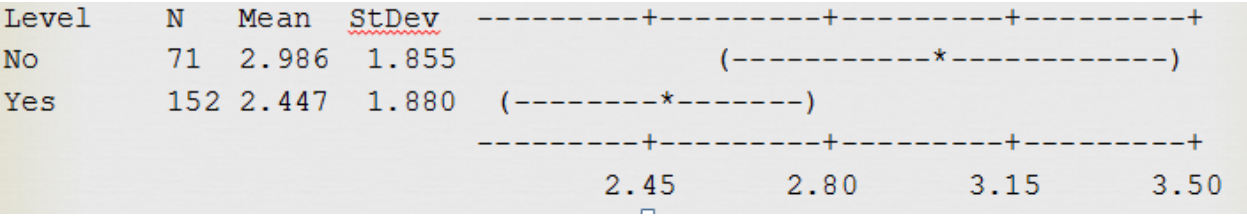
With a very conservative estimated response rate of 15%, the sample sizes for each class were 120, 108, 108, 102, and 18, respectively.

Once these figures were determined, we selected the sample. To do so, we used an Excel random number generator to generate pairs of random numbers (n_1, n_2), such that n_1 was the page number and n_2 was line number in C-book. That individual's email address was then recorded, and a master d-list composed. One of the biggest challenges we faced with respect to the survey creation was repetition. Sometimes after generating a pair of numbers, the corresponding name would either be a student we have already chosen or a student that wasn't supposed to be in our sample. In either case the solution was to generate new numbers and try again. Once the final distribution list was composed, we sent the survey out in email form. To provide an incentive to respond, we offered a $50 Amazon gift card to one individual who was selected at random from a list of those who voluntarily provided their Andrew IDs. In order to maintain confidentiality, providing this ID was purely optional, so as to preclude any possibility of linking names with responses. We had an initial problem with an under response from juniors, but upon investigating this further we discovered that the list for juniors was inherently flawed due to some minor miscommunication. We decided to send a new initial email to juniors to correct for this. One week after both initial invitations were sent, we sent out a reminder email encouraging people to respond. At the end of the survey, we were delighted to see a 50.4% response rate with 230 respondents.

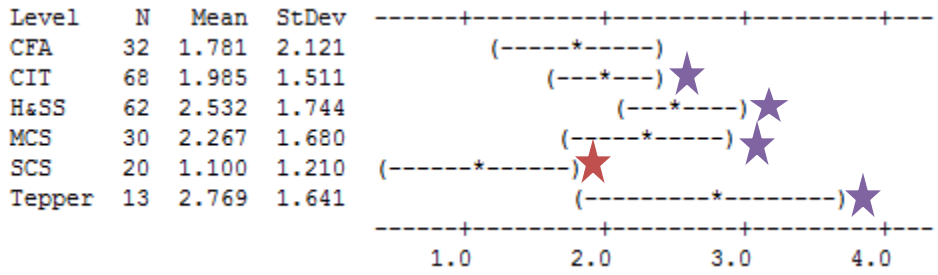
Results:

Our primary aim was to characterize student consumption of, and perception of, caffeine consumption on campus. The survey responses reached across a broad range of GPA, units taken, majors, and colleges. We are thus confident that the group is representative of the population. Confidence intervals and p values are calculated without the finite population correction. We did not decide to post stratify the results based on class because most of the variables we measured did not differ significantly by class on caffeine consumption (with the exception of fifth years and freshmen (p=.083)). The busyness score was calculated by summing involvement hours in extracurricular activities, classwork and job hours and then subtracting leisure and sleep hours. All negative scores were coded as 0. We looked in to post stratification based on GPA but could not find reliable population statistics for GPA and so determined that such a course was impossible.

Most of our analysis looks like this:

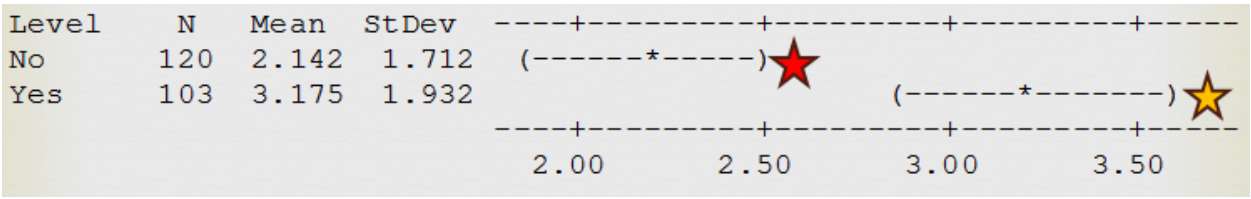


This is an ANOVA plot that shows no difference in caffeine consumption between those who believe that caffeine affects their health (the yes level) and those who believe that it has little or no impact (the no level)(p = 0.047) There were many other insignificant results. These included college and caffeine consumption, caffeine consumption and exercise hours, GPA and college, caffeine consumption and GPA, smoking habits and class, sleep hours and gender, and the result that students who believed that caffeine enhanced performance did not differ from their counterparts in GPA. Some significant results were that SCS students self-reported significantly fewer extracurricular hours than CIT, H&SS, MCS, and TSB students (purple stars) (p=0.012 for H&SS).

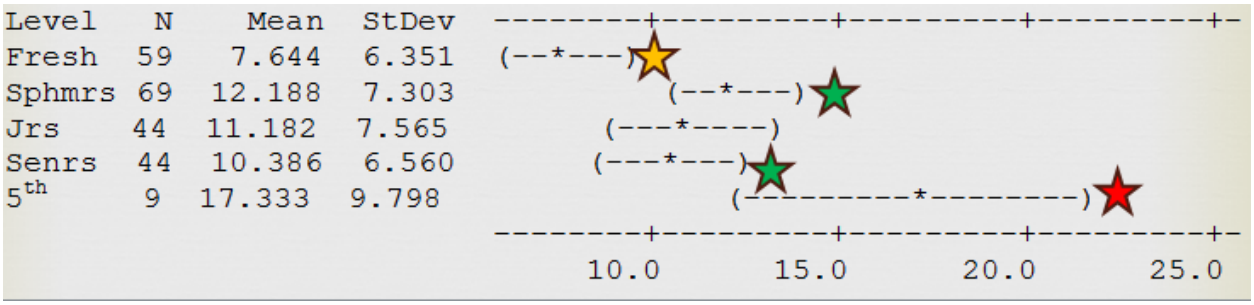


Also significant are the results that CFA students smoke significantly more than most other students (p = 0.027) and that CFA students report significantly less sleep than CIT and H&SS students (p=.017).

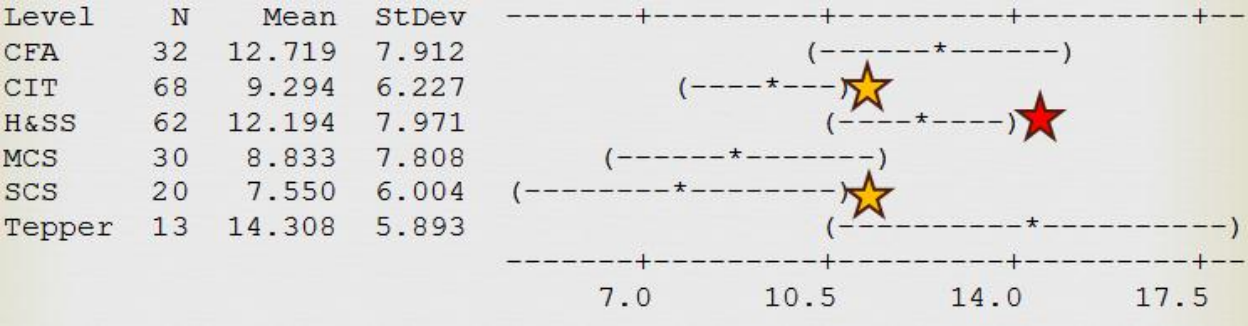
One of the largest differences that we found in caffeine consumption was between students who believed in the performance enhancement aspects of caffeine (consumed much more) and those who did not (consumed much less) (p<.001) as can be illustrated in the ANOVA plot below.



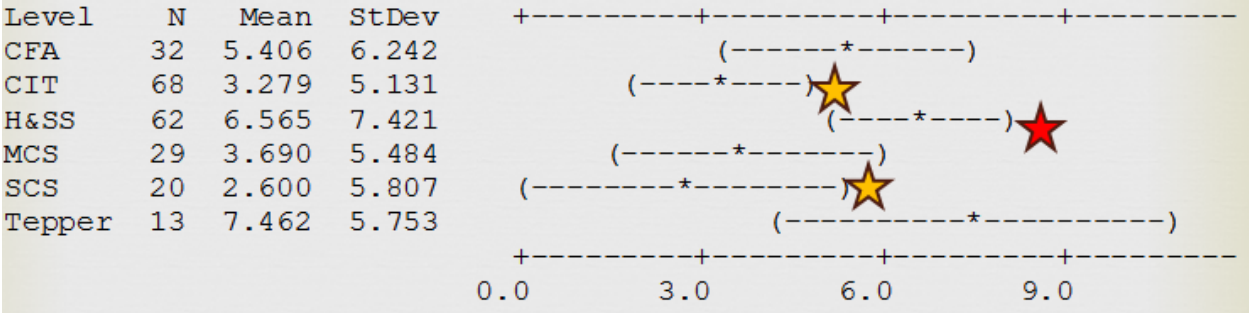
Freshmen are significantly less busy than sophomores (and probably all upperclassmen with the fpc) while seniors (and probably most underclassmen) were significantly less busy than 5th years (p <0.01) as shown below.



The busyness variable was spread out over the colleges as seen below:



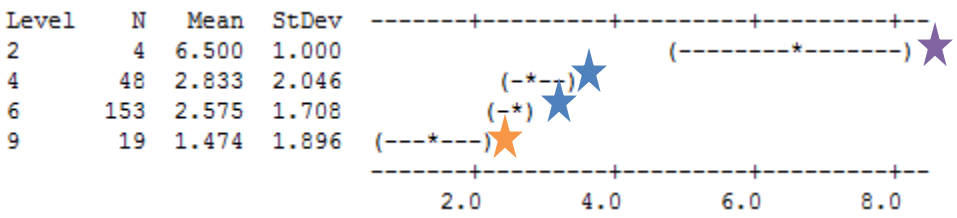
H&SS significantly beat out CIT and SCS students (p=.006), although CFA and Tepper were close seconds. One reason for this stereotype-defying result may be that students from those three schools reported more paid job hours than other colleges with H&SS significant (p=0.01).



Not surprisingly, students who drink caffeine (servings on the horizontal axis) tend to sleep less (hours on the vertical axis) (p<.001).

Students who consume more caffeine (vertical axis) also self-report more symptoms associated with caffeine addiction (horizontal axis) although this difference is insignificant until we compare students who consume no caffeine with those who consume more than seven servings a day(p=.007).

Symptoms of caffeine addiction are better predicted by sleep hours (p<.001), however, this may be due to the fact that many of the questions are sleep related.



After performing these ANOVA tests, we calculated a general regression model, excluding any variable pairs with high R^2 values and moderate Pearson correlation coefficients that could not be explained by the way that the variable was structured (i.e. categorical indicator variables). The model that we calculated is as follows:

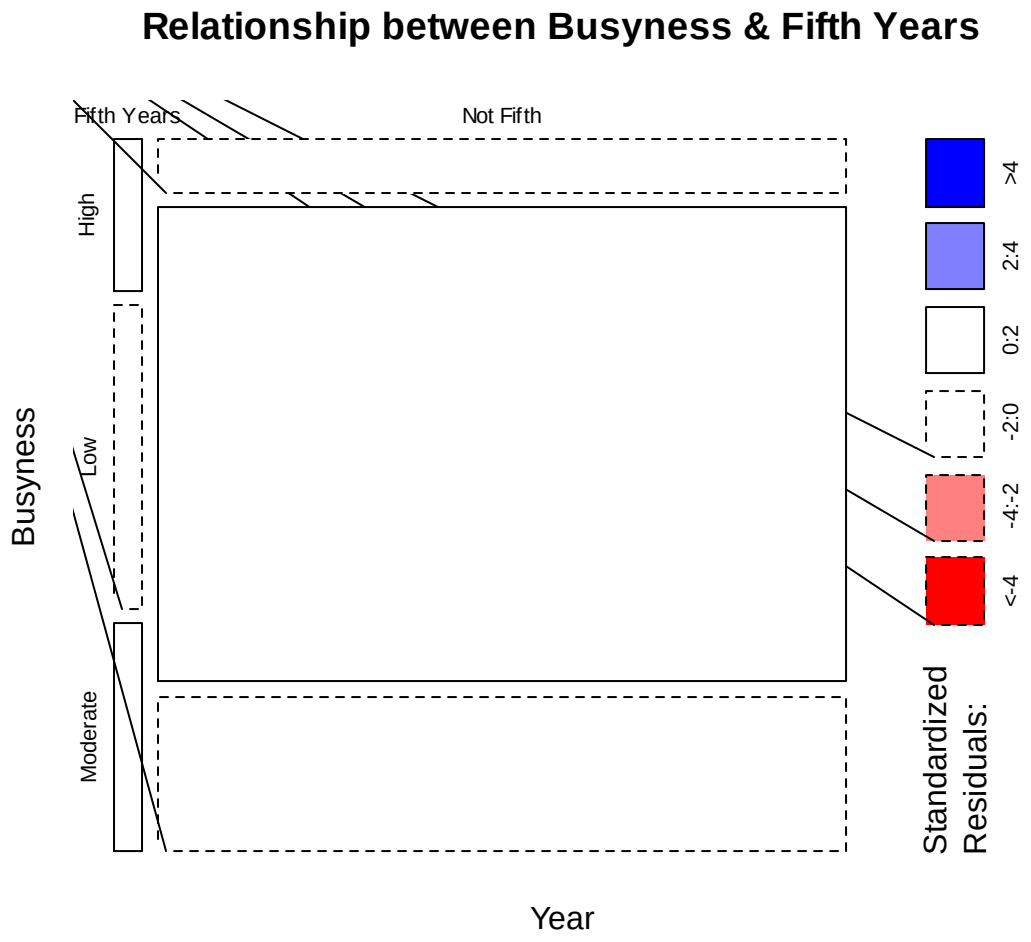
Caffeine Consumption = 2.31419 + 0.03118 * (Busyness) + 1.38030 * (Below Average Consumption) + 2.45278 * (Above Average Consumption) + -0.59294 * (Comfort Level) + -0.11757 * (Consumes to Stay Awake) + -1.9187 * (Does not consume caffeine) + 0.54911 * (Is a Fifth Year)

This model has an R^2 adjusted value of 40.1%, meaning that about 40% of the variability in caffeine consumption can be explained using this model. The p-values for these coefficients can be found in the table below. Although there are a few variables whose coefficients were not significant we included them under the suspicion that there was an interaction between those variables and at least one of the other variables in the model.

Variable	P-value
Busyness	p = 0.033
Below Average Consumption	p < 0.0009
Above Average Consumption	p < 0.0009
Comfort Level	p = 0.024
Consumes to Stay Awake	p = 0.593
Does not consume caffeine	p < 0.0009
Is a Fifth Year	p = 0.315

Furthermore, we wanted to test to see if there were any dependence between 5th Years and busyness score. The plot below is a Mosaic Plot which shows the dependent relationship amongst variables. In this case, we tested the hypothesis: *Busyness and Fifth Year are dependent*. We see that the resulting graph is all white (having no colors), representing an independent relationship between being a 5th year and busyness score. If these two variables were related, the blocks would appear colored (shade of the color would denote stronger dependent relationship). The size of the blocks tell us that there is a large

group of Non-Fifth Years who have *low* busyness while we see that there is a very small group of fifth-years who have *high* busyness scores.

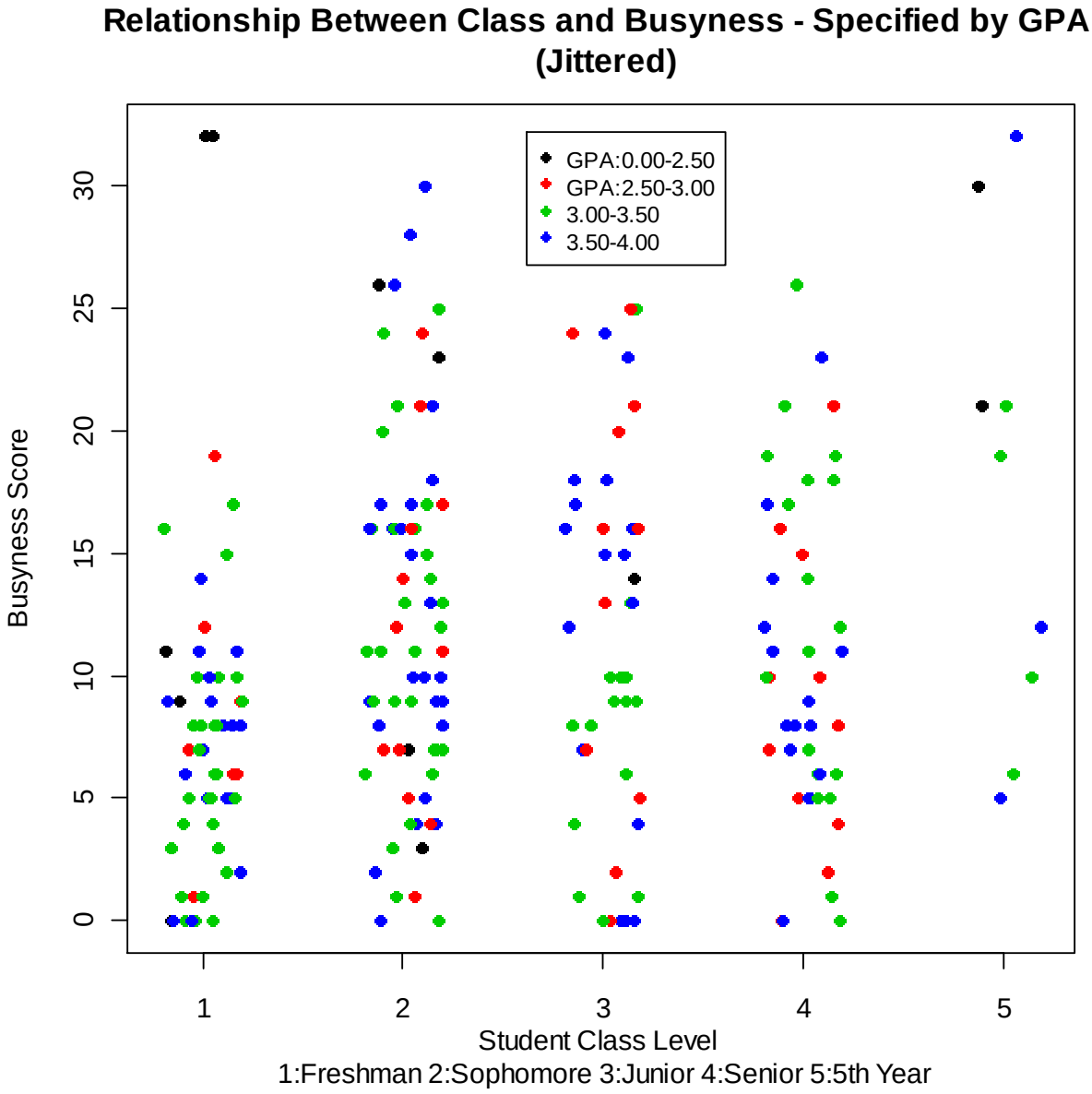


Although we do see some residuals that deviate away from the expected values conditioned on independence (the dotted lines), we still state that it is not enough evidence to suggest that the two variables, 5th years and Busyness, are dependent on each other. Because they are close to being independent, there is also no apparent trend that can be easily observed. The advantage of using the mosaic plot is that we are able to see whether or not the two defined variables are dependent or not, and this includes every single observation. The cut-offs of the busyness scores were pre-determined and were coded directly into R in the following way:

```
low.b=which(Busyness<=13)
med.b=which(Busyness>13 & Busyness<=21)
high.b=which(Busyness>21)
```

There were no major unexpected conclusions gathered from the data, except to say that these two variables seem to be independent.

In this graph, we are trying to portray the relationship amongst three variables: GPA, Busyness, and Class. We color code the different levels of GPA by the separate colors represented in the legend, and we divide the axes by Class Level and Busyness Scores (x and y respectively). We see that these are not very spread out horizontally because we are using a categorical variable (non-numeric), *class level*, and we see that there are fewer than 230 observations (the number of observations used for analysis). This is because we have overlapping points, which we will jitter in the next graph to better represent the observations. We see that there are much more green points, GPA: 3.00-3.50, and very little black points, GPA: 0.00-2.50. We can see one outlier in the in the freshman class where a student stated that he/she is extremely busy while maintaining a very low GPA.



To better observe the data, we jittered the *Class*. We are now able to observe almost all 230 observations. In terms of GPA, there seems to be a trend where the number of observations of the middle GPA group (3.00-3.50) significantly decreases in the junior year. Rather, most of the GPAs become either very high or very low, with barely any in between. The variation of the busyness level is much smaller in the freshman year and is limited to being beneath 20, excluding the outliers. Sophomores, juniors, and seniors show a similar range and variance of busyness, which can be expected since the rigor of the course tends to pick up during these years. 5th years have the largest variance, which may be due to the small number of observations. Jittering the class gives us an idea of the time periods in between the categorized classes and creates a better sense of what a random 'noise' in the data might look like.

Discussion:

Overall, we were pleased with the survey administration. The very high response rate is, we believe, a tribute to the Amazon gift card incentive, which proved to be very effective.
The survey responses did answer our initial research questions which we set out to address, although we did have some very surprising results. Contrary to our expectation, we did not see any significant difference among underclassmen (all but 5th years) in their caffeine consumption. Though a significant relationship does exist between busyness and caffeine consumption, it was not as large of an effect as we had anticipated. We had also expected to see a more clear relationship between health concerns and caffeine consumption, but were very surprised to see that there was no significant caffeine consumption difference between those who felt that caffeine consumption was bad for them and those who thought it was healthy.

The insignificant interaction between belief in the health detriments of caffeine and actual caffeine consumption is interesting because it shows that consciousness of the health risks of caffeine does not significantly predict smaller consumption levels. Most of the other insignificant results are uninteresting as they merely confirm the general perception that such variables are unconnected. Two significant results that confirm school stereotypes are that SCS students have significantly fewer extracurricular hours and CFA students get less sleep and are more likely to smoke. Some results that disprove college stereotypes are that H&SS, Tepper, and CFA students are busier than SCS and CIT students, although this may be due to their tendency to report more paid job hours. Our busyness variable shows an increased busyness as year increases, which mirrors the increasing workload of students and extracurricular involvement, so we are confident in its validity as a measure. The decreases in sleep hours as we increase caffeine consumption and the increase in caffeine addiction symptoms as caffeine consumption increased were both indicative of the internal validity of our caffeine consumption questions and since they follow regular patterns we are confident that our questions were unambiguous. It is interesting though that the difference in symptoms reported is only significant between students who consume no caffeine and those who consume more than seven servings a day, although this may be due to the fact that many of the symptoms can also be caused be lack of sleep, which students who do not consume caffeine may still experience. The coefficient for busyness from our regression model (.03118) can be interpreted as the increase in caffeine consumption per one point increase in busyness score, on average, with all other modeled factors held constant. Because busyness ranged from 0 to 32 we can expect the busiest students to drink about 1 more serving of caffeine per day, on average, than their less busy counterparts.

One potential weakness of our study is a sort of self-selection bias. One of the primary variables we were interested in was student busyness, but by definition, students who are busier will have less time available to fill out surveys for other people. Thus there could be systemic bias in our results which will cause us to actually underestimate how busy students are on average because we are not getting responses from the busiest students. Future iterations of this survey are encouraged to try to think of a way to better capture these responses and therefore achieve more accurate results. Another is the highly variable group of 5^{th} years, some of which are highly driven students pursuing accelerated masters degrees, and some of which are struggling students who are trying to complete their undergraduate degrees. Future groups are encouraged to investigate this interesting yet poorly understood group of students.

The survey did achieve its stated goal, and the question structure was precise, objective, and consistent. One major strength of the survey was the use of the busyness score tactic, with which we were able to gain a very accurate perspective on exactly how busy students are (rather than asking a subjective question asking students how busy they felt they were). Calculating our own busyness score ensured objectivity and consistency throughout the sample pool.

The take home message of the survey is that generally, students do consume more caffeine if they believe it will enhance their performance. Most people believe they consume a below average amount of caffeine (a possible egocentric bias), although there are not significant correlations with class year (except 5^{th} years) or GPA. Future groups are encouraged to try to address the self-selection bias of busy students and to investigate further the 5^{th} year students.

Appendix A: Full Questionnaire:

1. What is your year at CMU?
    a. Freshman
    b. Sophomore
    c. Junior
    d. Senior
    e. Fifth year undergraduate student
2. What college is your primary major in? (For interdisciplinary, select both)
    a. CIT
    b. H&SS
    c. CFA
    d. SCS
    e. MCS
    f. Heinz
    g. Tepper
3. What is your GPA?
    a. 0.0-2.5
    b. 2.6-3.0
    c. 3.0-3.5
    d. 3.6-4.0
4. What is your gender?
    a. Male
    b. Female

5. How many units are you taking this semester?
    a. 36 or less
    b. 37-45
    c. 45-55
    d. More than 55
6. Do you have any additional majors or minors?
    a. Yes
    b. No
7. If so, which statement best describes your situation?
    a. I have a minor
    b. I have more than one minor
    c. I have an additional major
    d. I have an additional major and an additional minor
    e. I have more than one additional major
    f. I have more than one additional major and more than one minor
8. About how many hours a day, on average, do you spend on extracurricular activities such as clubs, student organizations, ROTC, or inter-mural sports (this includes attending and preparing for meetings or events)?
    a. None
    b. Less than 2 hours per day
    c. 2-4 hours per day
    d. 4-6 hours per day

    e. More than 6 hours per day
9. How much personal leisure time do you have in a given day? This includes all waking hours spent on activities you consider optional (i.e., NOT class, work, athletics, extracurricular obligations).
    a. None
    b. Less than 2 hours per day
    c. 2-4 hours per day
    d. 4-6 hours per day
    e. More than 6 hours per day
10. How many hours outside of class per day, on average, do you spend doing classwork?
    a. None
    b. Less than 2 hours per day
    c. 2-4 hours per day
    d. 4-6 hours per day
    e. More than 6 hours per day
11. How many hours do you sleep on an average week night?
    a. Less than 3 hours
    b. 3-5 hours
    c. 5-8 hours
    d. More than 8 hours
12. Do you have a part-time job (employed on or off campus)?
    a. Yes; I work for less than 5 hours per week.
    b. Yes; I work for less than 10 hours per week.
    c. Yes; I work for less than 15 hours per week.
    d. Yes; I work for more than 15 hours per week.
    e. No; I do not have a job.

13. How often do you exercise in an average week?
    a. None
    b. Once a week
    c. 2-4 times per week
    d. More than 4 times per week
14. What types of caffeine beverages do you consume? Select all which are relevant.
    a. Coffee
    b. Tea
    c. Espresso
    d. Caffeine tablets
    e. Redbull/energy drinks
    f. Soda
    g. Chocolate
    h. I do not consume caffeine.
    i. Other _____
15. How much caffeine do you consume on a daily basis, where 1 serving is one item (eg. One cup of coffee, one espresso shot, one can of soda, one bar of chocolate, etc..)
    a. None
    b. Less than 2
    c. 2-4
    d. 4-6

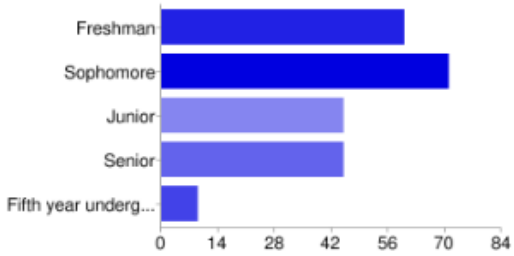
    e. More than 6
16. What is your primary motivation for consuming caffeine?
    a. Pleasure/Leisure
    b. To stay awake/maintain energy
    c. Partying
    d. Other _____
17. Are you consuming more or less caffeine than you are comfortable with?
    a. I am consuming more caffeine than I am comfortable with.
    b. I am consuming a comfortable amount of caffeine.
    c. I am drinking less caffeine than I am comfortable with
18. Do you feel that your personal caffeine consumption is more or less than the average consumption of your undergraduate peers at Carnegie Mellon?
    a. I believe I consume an above average amount of caffeine.
    b. I believe I consume an average amount of caffeine.
    c. I believe I consume a below average amount of caffeine.
19. Do you believe that consuming more caffeine positively impacts your performance in school work, employment, athletics, or extracurricular activities?
    a. Yes
    b. No
20. Do you believe that consuming more caffeine negatively impacts your health?
    a. Yes
    b. No
21. Do you experience any of the following on a regular basis (on most days)? Choose all that apply.
    a. Insomnia
    b. Nervousness/anxiety throughout the day
    c. Restlessness
    d. Irritability
    e. Nausea/gastrointestinal problems
    f. Fast or irregular heartbeat
    g. Muscle Tremors
    h. Headaches
    i. I do not experience any of these symptoms.
22. Do you smoke?
    a. Yes
    b. No
    c. I do not wish to disclose
23. How frequently do you consume alcohol?
    a. None
    b. Less than twice a week
    c. 2-4 times per week
    d. More than 4 times per week

Appendix B: Summary Statistics for Each Survey Question

230 responses

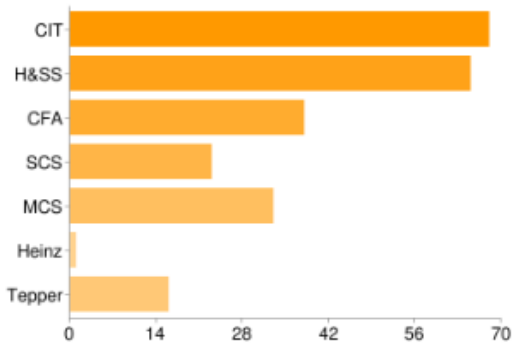
Summary See complete responses

What is your year at CMU?



Freshman	60	26%
Sophomore	71	31%
Junior	45	20%
Senior	45	20%
Fifth year undergraduate student	9	4%

What college is your primary major in? (For interdisciplinary, select both)

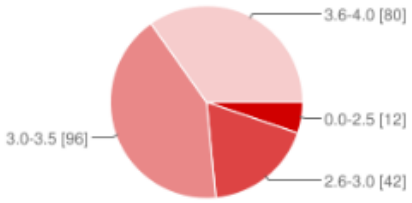


CIT	68	30%
H&SS	65	28%
CFA	38	17%
SCS	23	10%
MCS	33	14%
Heinz	1	0%
Tepper	16	7%

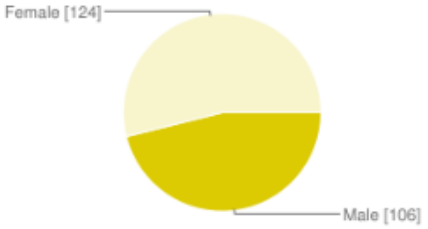
People may select more than one checkbox, so percentages may add up to more than 100%.

What is your GPA?

0.0-2.5	12	5%
2.6-3.0	42	18%
3.0-3.5	96	42%
3.6-4.0	80	35%

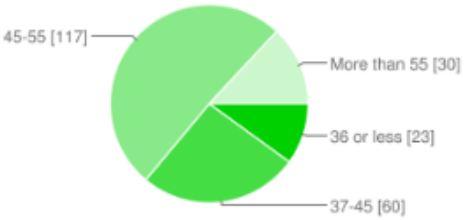


What is your gender?



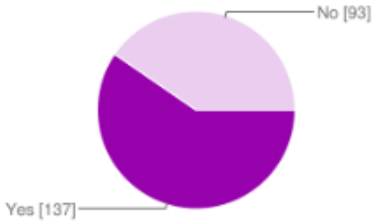
Male	**106**	46%
Female	**124**	54%

How many units are you taking this semester?



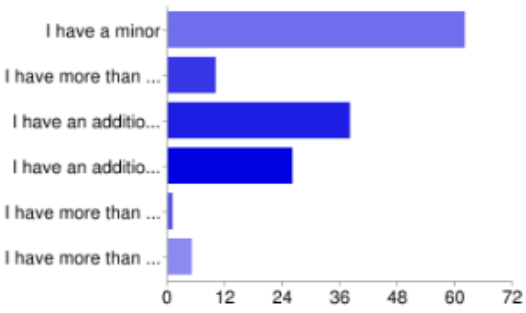
36 or less	**23**	10%
37-45	**60**	26%
45-55	**117**	51%
More than 55	**30**	13%

Do you have any additional majors or minors?



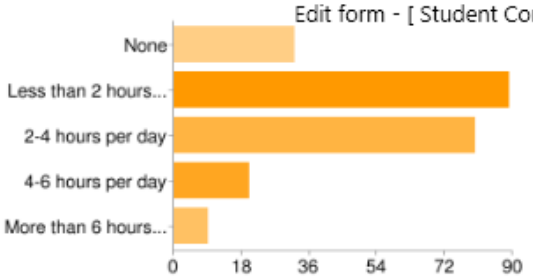
Yes	**137**	60%
No	**93**	40%

If so, which statement best describes your situation?



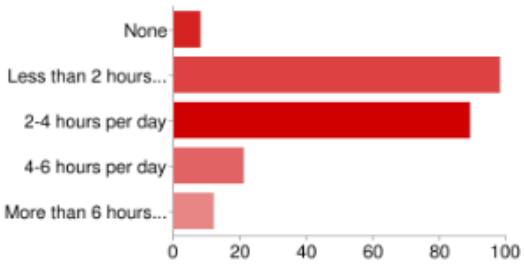
I have a minor	**62**	44
I have more than one minor	**10**	7
I have an additional major	**38**	27
I have an additional major and an additional minor	**26**	18
I have more than one additional major	**1**	1
I have more than one additional major and more than one minor	**5**	4

About how many hours a day, on average, do you spend on extracurricular activities such as clubs, student organizations, ROTC, or inter-mural sports (this includes attending and preparing for meetings or events)?



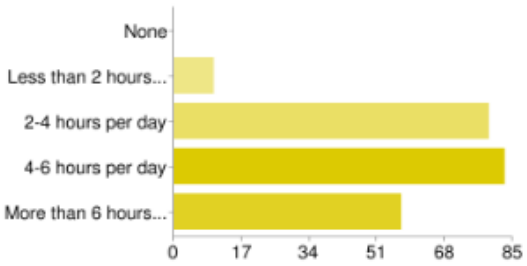
Less than 2 hours per day	89	39%
2-4 hours per day	80	35%
4-6 hours per day	20	9%
More than 6 hours per day	9	4%

How much personal leisure time do you have in a given day? This includes all waking hours spent on activities you consider optional (i.e., NOT class, work, athletics, extracurricular obligations).



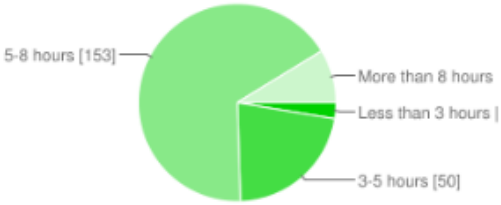
None	8	4%
Less than 2 hours per day	98	43%
2-4 hours per day	89	39%
4-6 hours per day	21	9%
More than 6 hours per day	12	5%

How many hours outside of class per day, on average, do you spend doing classwork?



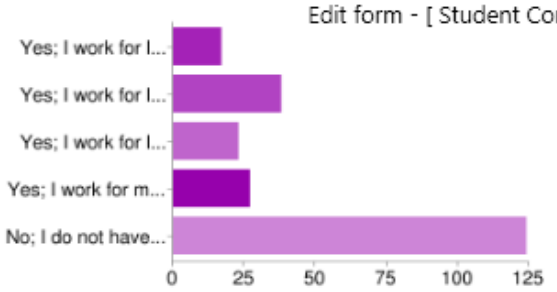
None	0	0%
Less than 2 hours per day	10	4%
2-4 hours per day	79	34%
4-6 hours per day	83	36%
More than 6 hours per day	57	25%

How many hours do you sleep on an average week night?



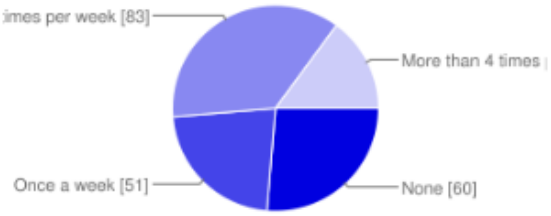
Less than 3 hours	6	3%
3-5 hours	50	22%
5-8 hours	153	67%
More than 8 hours	20	9%

Do you have a part-time job (employed on or off campus)?



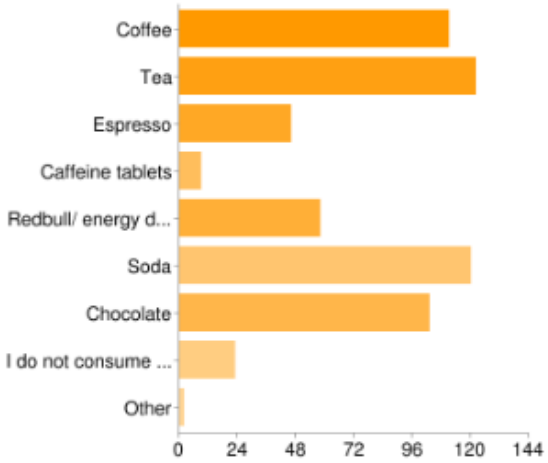
Yes; I work for less than 5 hours per week.	17	7%
Yes; I work for less than 10 hours per week.	38	17%
Yes; I work for less than 15 hours per week.	23	10%
Yes; I work for more than 15 hours per week.	27	12%
No; I do not have a job.	124	54%

How often do you exercise in an average week?



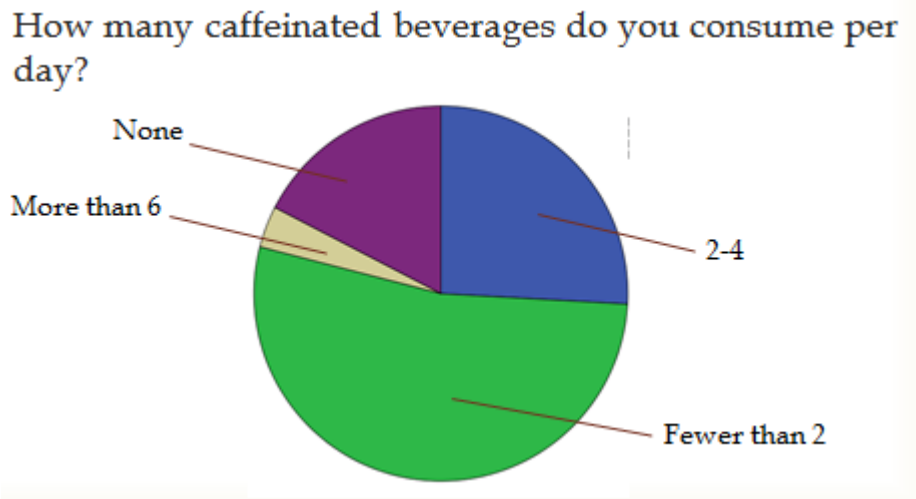
None	60	26%
Once a week	51	22%
2-4 times per week	83	36%
More than 4 times per week	34	15%

What types of caffeine beverages do you consume? Select all which are relevant.

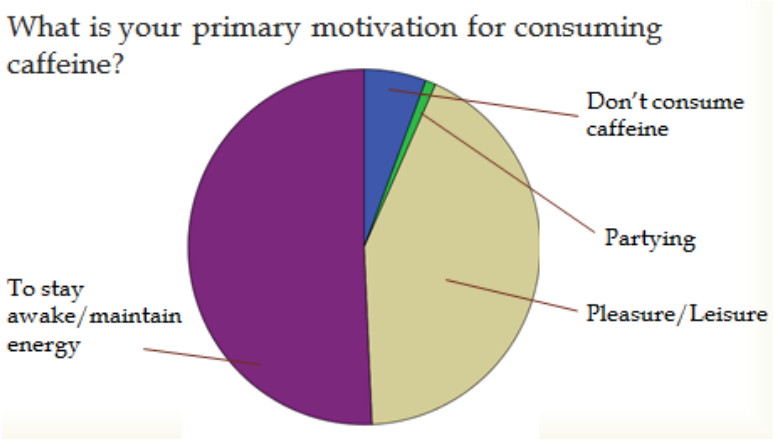


Coffee	111	48%
Tea	122	53%
Espresso	46	20%
Caffeine tablets	9	4%
Redbull/ energy drinks	58	25%
Soda	120	52%
Chocolate	103	45%
I do not consume caffeine	23	10%
Other	2	1%

People may select more than one checkbox, so percentages may add up to more than 100%.

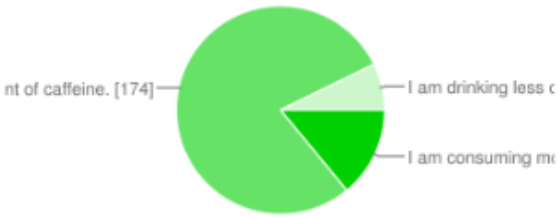


Note: The spreadsheet incorrectly coded some of these data as dates. We therefore corrected that and manually created this graph.



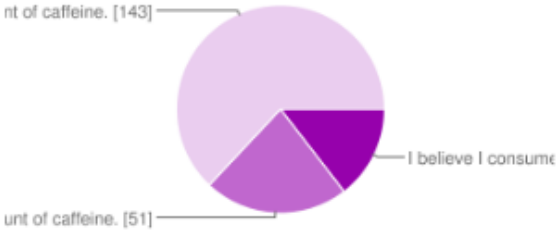
Note: The spreadsheet incorrectly coded some of these data as dates. We therefore corrected that and manually created this graph.

Are you consuming more or less caffeine than you are comfortable with?



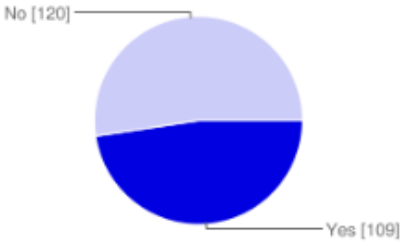
I am consuming more caffeine than I am comfortable with.	31	14%
I am consuming a comfortable amount of caffeine.	174	79%
I am drinking less caffeine than I am comfortable with	16	7%

Do you feel that your personal caffeine consumption is more or less than the average consumption of your undergraduate peers at Carnegie Mellon?



I believe I consume an above average amount of caffeine.	33	15%
I believe I consume an average amount of caffeine.	51	22%
I believe I consume a below average amount of caffeine.	143	63%

Do you believe that consuming more caffeine positively impacts your performance in school work, employment, athletics, or extracurricular activities?

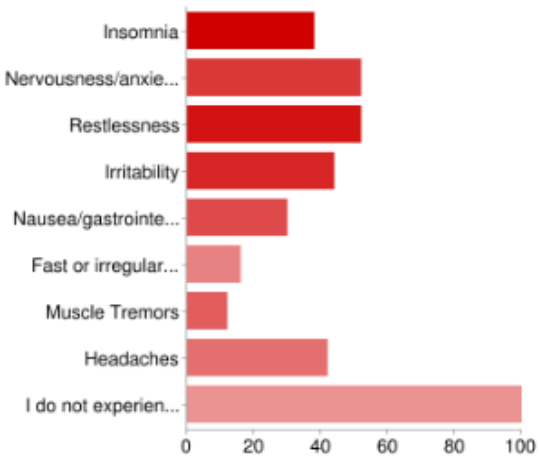


Yes	109	48%
No	120	52%

Do you believe that consuming more caffeine negatively impacts your health?

Yes	156	68%
No	73	32%

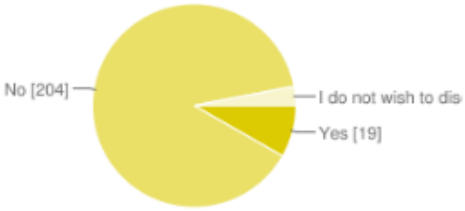
Do you experience any of the following on a regular basis (on most days)? Choose all that apply.



Insomnia	38	17%
Nervousness/anxiety throughout the day	52	24%
Restlessness	52	24%
Irritability	44	20%
Nausea/gastrointestinal problems	30	14%
Fast or irregular heartbeat	16	7%
Muscle Tremors	12	5%
Headaches	42	19%
I do not experience any of these symptoms.	100	45%

People may select more than one checkbox, so percentages may add up to more than 100%.

Do you smoke?



Yes	19	8%
No	204	89%
I do not wish to disclose	7	3%

Appendix C: Informed Consent Statement:

Confidentiality Policy:

This survey tests student caffeine consumption at Carnegie Mellon University, and is a course requirement for course 36-303: Sampling, Surveys, and Society, taught by Professor Brian Junker. This research project guarantees respondent confidentiality. We conform to all aspects of the Ethics Code of the American Association of Public Opinion Researchers. Most importantly, we guarantee that we will honor clause D-2 of the Ethics Code, namely "Unless the respondent waives confidentiality for specified uses, we shall hold as privileged and confidential all information that might identify a respondent with his or her responses. We shall also not disclose or use the names of respondents for non-research purposes unless the respondent grants us permission to do so."

All data will be used in a form that will make it impossible to determine the identity of the individual responses. That is, the survey responses will not be integrated, analyzed, or reported in any way in which the confidentiality of the survey responses is not absolutely guaranteed. Access to raw data will be tightly restricted to only those individuals directly involved in data analysis. The survey administrators will retain the sole ownership of all raw data.

The final survey report will be made available by request to any interested party. Please email yonggyuc@andrew.cmu.edu or the course advisor, Brian Junker, at brian@stat.cmu.edu if you have any questions about this policy.

Appendix D: Survey Email Distributed by Email

Dear Student,

You have been randomly selected to participate in a campus-wide survey on caffeine usage among Carnegie Mellon undergraduate students. This very brief survey attempts to assess how much caffeine, and what types, students consume. Please take just a minute to complete the survey--all students who complete the survey will be entered into a raffle to win a $50 Amazon gift card!

If you have trouble viewing or submitting this form, you can fill it out online:
https://spreadsheets.google.com/viewform?formkey=dHY3cmhvZUVkQ2hQX205RjZVUl92bEE6MA

Raffle Entry;
If you would like to be entered into the raffle to win the $50 Amazon gift card, please click the survey link below AFTER YOU SUBMIT your survey responses. We guarantee the confidentiality of your responses, and listing your andrew ID here will not in any way link your name to your survey responses. Click
here: https://spreadsheets0.google.com/viewform?hl=en&hl=en&formkey=dERubDhsT2lpWGhMeUhWTEpCU2RGOUE6MQ#gid=0

References

CMU Admissions (2011). *Carnegie Mellon Factbook*, Enrollment Statistics <http://www.cmu.edu/ira/factbook/pdf/facts2011/4_enrollment-final-as-of-2_24_11.pdf>

Erst, S (2008). The Caffeine Addition Problem. *The Standard Online.* http://media.www.the-standard.org/media/storage/paper1059/news/2008/02/05/Features/The-Caffeine.Addiction.Problem-3189385.shtml

Mayo Clinic Staff (2011). Caffeine: How much is too much? *Mayo Clinic Health Information*. http://www.mayoclinic.com/health/caffeine/NU00600