Internet Piracy:
CMU Undergraduate Behavior & Opinions

James Bogie | Ben Gorman | Chelsea Grindle
Jong Woo Lee | George Nardi | George Volichenko

May 4, 2012

36-303
Sampling, Surveys, & Society
Professor Brian Junker

Section 1: Introduction

In the age of the internet, democratized information poses a threat to intellectual property (IP). In the U.S., intellectual property rights are protected via trademarks, trade secrets, patents, and copyrights. Copyrights protect creative works like music, books, and movies. Given the universal popularity of these goods and the ease with which they can be digitally replicated, internet-based information trafficking has notoriously undermined copyrights. According to an August 2011 study published by The American Assembly (Columbia University public affairs forum), 46% of adults have bought, copied, or downloaded unauthorized music, TV shows, or movies. Among 18-29 year olds, 70% have acquired music or video files.

Advocates of stricter copyright enforcement online sought more stringent protection of IP by proposing the SOPA (Stop Online Piracy Act) and PIPA (Protect Intellectual Property Act) legislation in October and May of 2011, respectively. When the bills first made headlines, they garnered a good deal of support from lawmakers, but a flood of both grassroots (social media, public protests) and institutional (Wikipedia, Google) opposition quickly attenuated that support. Online communities and sites around the web observed a "blackout" to increase awareness of the proposed bills. Google hired at least 15 lobbying firms to fight the legislation, and Mozilla put a link to a petition on its homepage with the warning "Congress is trying to censor the internet" (Wyatt). We're curious to know whether opposition to the bills (at least in our undergraduate community) stems from skepticism of bureaucrats tampering with the internet, or if the backlash to regulation comes from internet users that have grown accustomed to pirating content. Regulating IP online is a complex legal issue, with strong opinions on both sides. Lawrence Tribe, a law professor at Harvard, opposes SOPA and even considers it illegal because some of the would-be-removed content is constitutionally protected free-speech (Breen).

As the American Assembly survey shows, internet piracy is a widespread behavior among young people. However, the *extent* to which people participate in these behaviors is still unknown. Furthermore, it remains unanswered how young people view these activities ethically and legally, and whether or not their views are consistent with their actual behavior. Our survey intends to measure the prevalence of internet piracy prevalence on campus, assess undergraduate

opinions on internet regulation, and garner information about online pirating behaviors and methods. We want to know how often students pirate, what types of content they download or stream, and what types of websites they use. Respondents who claim to not engage in piracy are also interesting; we would like to learn more about their attitudes toward piracy, particularly by finding out what keeps them from illegally viewing movies, shows, and music. We collected demographic information from respondents to both mitigate potentially confounding results and also to learn about behavioral differences between demographics.

Section 2: Methods

Target population

The target population for this survey was the undergraduate student body at Carnegie Mellon University (Pittsburgh, Pennsylvania). We assumed a slight difference between the target population and the sampling frame results from the inclusion in the student directory of those who have dropped out, transferred, or otherwise taken a leave of absence. We considered contacting students via Facebook to mitigate potential non nonresponse from e-mail solicitations, but after considering the sufficient sample size and relatively low margin of error obtained from e-mail solicitations alone, we decided to send out surveys only through e-mail. We anticipated a large nonresponse error: historically, most people who receive survey solicitations via e-mail do not participate. However, simply contacting enough people to take the survey eventually yielded an adequate number of responses for statistical analysis. While there is the potential for measurement error in our results due to a respondent misinterpreting a question, we emphasized unequivocal language in the survey, and also provided clarifying definitions when necessary, e.g., "Have you ever streamed pirated content? Example: Stream a TV show from Megavideo/SideReel".

Sampling Frame

We drew our sample from Carnegie Mellon undergraduate students, using the Carnegie Mellon C-Book (student directory) as the sampling frame. We generated a random sample from the set of page numbers of the C-Book with *R*. Then, we recorded all the undergraduate email addresses on those pages for survey solicitation. In the instance that the surnames on a given C-Book page

were homogeneous enough, we excluded that page in order to avoid over-representing a certain demographic, e.g., Kim, a common Korean surname, represents 96% of the names on page 58.

Sample Design

The format used was a self-administered, web-based survey with email as the contact method. The e-mail solicitation is the sole instance of "interviewer involvement" or degree of interaction with the respondent. Results from the survey were tabulated to a third-party host site, a Google Docs-based survey generator. To safeguard privacy and generate honest answers, we emphasized that responses are strictly confidential. If respondents wanted a chance at winning a prize, providing an email address was optional. The survey was text-only (no audio or video). The technological requirements of the survey were limited: respondents only needed to fill out the survey online. While it is possible that programming error resulted from variable internet connections and computer capabilities, we anticipate this potential error to be relatively minor given that CMU's academic environment requires most of its students to have adequate computing technology and internet access.

Sample Size

To approximate an appropriate sample size we considered one of the many important questions on our survey: "Have you ever streamed pirated content?" Ideally, we wanted the margin of error to be as small as possible. Using 0.01 as the margin of error, the required sample size would have to be 3,760. Since time constraints do not allow such a large sample size, we tried a margin of error of 0.05 instead. This accuracy level required a much more realistic sample size of 362.

Assuming a 25% response rate, 1448 undergraduates would have to be solicited to generate 362 responses. Once we actually started sampling Andrew IDs from the C-Book, we realized that it is a rather time-consuming process. The Andrew IDs were inputted to a spreadsheet, which was turned into a csv file for further processing. Using a simple 'for' loop in R, every ID was concatenated with "@andrew.cmu.edu". The first wave of e-mails included a subset of only 200, as we wanted to get an approximation of our actual response rate. The initial response rate was 15.81% (31/196) in the first 24 hours. The full sample of e-mails consisted of 1081 elements, 24 of which turned out to be misspelled. It was very convenient that the e-mail client indicated the

addresses that were misspelled during the C-book stage. We simply removed those e-mails from the sample. After we contacted the rest of the sample, 183 students responded to the survey, making the response rate 17% (183/1057). To increase the number of respondents we sent a courteous reminder to the same sample of 1057 undergraduate students. That resulted in 57 more responses, bringing the final response rate up to 23% with 240 total responses.

We then recomputed the margin of error for the same question: "Have you ever streamed pirated content?". Using our final number of responses and the actual proportion of "yes" responses (instead of p = 0.5):

$n = N * n0/(N + n0)$ $\Longrightarrow$ $n0 = (N * n)/(N - n) = 6178*240/(6178-240) = 249.7$

$MOE = z(0.0025) * sqrt(p*(1-p) / n0) = 1.96 * sqrt(0.758 * 0.242 / 249.7) = 0.053$

That means that the margin of error is actually very close to the one we forecasted. Initially using a conservative estimate of p=0.5 compensated for a smaller sample size in the calculation.

Survey Setup

Below are a few example questions from the different sections of our survey:

Background questions (SOPA/PIPA attitudes)

Example) Q.1 Have you heard of the recently proposed SOPA/PIPA bills?
- Yes or No

(Regardless of answer, we then provide summary information on both bills before the respondent completes the rest of the survey)

Personal piracy behavior questions

Example) Q.6 Have you ever streamed pirated content?

Example: Stream a TV show from Megavideo/SideReel.

- Yes or No (next question conditional on response)

Potential scenario questions

Example) Q.15 You download this movie from dtella (or another peer-to-peer file-sharing network) and watch it on your computer.

Do you consider this to be (check all that apply):
- Acceptable
- Illegal
- Unethical
- No opinion on the ethics or legality of this action

Demographic questions

Example) Q.26 What school are you in? (If you are in multiple schools, please select the school of your primary major.)

- CFA, MCS, CIT, HSS, Tepper, SCS

Post-survey processing

In terms of imputing any missing responses, we found that the amount of missing values was so low that it did not seem to make any kind of impact on our results. The highest number of missing values for any one variable was 3; most questions had 1 or zero missing values. The nature of the questions with missing responses also suggests that respondents skipped it by accident as opposed to intended non-response. We analyzed the relationship between missingness and the responses to other questions and did not find any clear relationships, so we concluded that the values are missing completely at random and omitting them would be an appropriate solution.

To determine whether post-stratification weighting was necessary, we examined the demographics of our respondents (relevant tables below). 57% of the respondents were male, and 42% were female. We found that 3 respondents did not include their gender (hence our percentages do not add to 100%). The sample gender proportion almost matches that of the overall undergraduate student body at CMU perfectly (both percentages off by less than half a percent), so we consider weighting for gender to be unnecessary.

Class Year	Sample Proportion	Population Proportion (Sample Proportion Needed)	Difference (Sample - Population)
Freshman	27.7%	25.5%	2.2%
Sophomore	23.1%	25.5%	-2.4%
Junior	19.7%	24.1%	-4.4%
Senior	26.9%	22.4%	4.5%
5th Year Senior	2.5%	2.4%	.1%

Gender	Sample Proportion	Population Proportion (Sample Proportion Needed)	Difference (Sample - Population)
Male	57.6%	58%	-.4%
Female	42.1%	42%	.1%

School	Sample Proportion	Population Proportion (Sample Proportion Needed)	Difference (Sample - Population)
CFA	11.7%	16.6%	-4.9%
CIT	32.9%	29.7%	3.2%
HSS	22.1%	19.9%	2.2%
MCS	14.6%	12.3%	2.3%
SCS	15.0%	9.8%	5.2%
Tepper	3.8%	6.1%	-2.3%

In comparison to the true population (CMU undergraduate student body), sophomores and juniors were underrepresented in our sample, while there were too many freshmen and seniors. In order to get our sample to better represent the population, we needed to incorporate post-stratified weights.

Looking at the school variable, we saw that the sample proportions were very close to the population proportions. We saw that our sample proportions for CFA and Tepper were below the population proportions, while CIT, HSS, MCS, and SCS were above the population proportions. The largest difference between sample and population proportions was 5.2% for SCS.

Given that we do not have a response from every demographic subset, we are forced to limit how many strata we choose to weight. For instance, while we do have responses from most class years in Tepper, we do not have responses from any freshman. The gaps in our demographic become even more drastic when we consider race. As an example, we do not have responses from any Hispanic or African-American juniors. If we wanted to place weights on multiple strata, we would need to impute values for the missing demographic components. Given the number of levels we had for our ethnicity and school variables, imputation would be a very difficult task, and none of the traditional methods seemed to offer great benefit to our results. In order to determine if imputation was necessary, we utilized graphical techniques to test whether there was any dependence between race or school and the responses to a few of our survey questions. We found convincing evidence that there was no dependence between the variables (see mosaic plots, Appendix Figures 2 & 3). Therefore, we are comfortable in choosing to only weight *class year*, the demographic variable for which we see the largest deviation from the target population. These weights will allow us to fix disproportionate sampling of post strata and disproportionate nonresponse across post strata.

A surprising find was that 34% of the respondents were Democrats, whereas only 9% reported themselves as Republican. We do not have any data on the political affiliations of the true population, so weighting was not an option.

Weighting

As previously mentioned, we decided to use post-stratification weights based on class year. We found the stratification weights based off class year and applied them to several different variables. The variables for which we applied weights were those that we chose to convert into numeric variables (essentially, ordinal categorical). Both the streaming and downloading variables were yes or no questions that we changed into a 0 or 1 variable. A 1 meant that the individual had downloaded and a 0 meant that they had not. Using this we could find the total mean, which is interpreted to be the total proportion of the sample that said they have streamed or downloaded. We then broke this down into the different strata and found each individual mean and variance using the Taylor series and jackknife methods. We also turned the streaming

frequency and downloading frequency variables into numeric variables as follows:

Only once	Few times a year	Few times a month	Weekly	Daily
1	2	3	4	5

Any person that chose not to answer a question we left out of this numeric conversion. When doing the calculations for these averages, we first found $\bar{y}w=(iyiwi)/ iwi$. When calculating the taylor series method for variance we used the following calculation: $VarTS(\bar{y}w)=1/(iwi)2 [Var(iwiyi)-2\bar{y}wCov(iwiyi,iwi)+\bar{y}w2Var(iwi)]$. After this variance was found we then calculated the sample mean and variance treating the survey as a simple random survey so that we could find the DEFF. In every case the DEFF was less than one, meaning that it was worthwhile to stratify on class.

While calculating the weights for the classes, we needed the total number of students from each class from the population, which we found in the CMU factbook. We found Wh by taking the number of students in each class over the total. We also had to find the fraction of people that took the survey in each of the classes. The weights are calculated by taking the population proportion divided by the sample proportion.

Strata	Nh	Wh	nh	fh	nh/n	Weights
Freshman	1559	0.2549	66	0.0423	0.2773	0.9194
Sophomore	1560.8	0.2552	55	0.0352	0.2311	1.104
Junior	1475	0.2441	47	0.0318	0.1974	1.221
Senior	1371	0.2242	64	0.0466	0.2689	0.8338
5th Year	148.7	0.0243	6	0.0403	0.0252	0.9646
Total	6114.5	1	238	--	1	5.0440

Section 3: Results

We have three main research questions:

(1) To what extent do students engage in piracy?
(2) How aware are students of the SOPA and PIPA bills?
(3) What are students' opinions on internet piracy and copyright enforcement policy?

Based on the 240 responses we gathered over the course of 21 days in April, our findings are as follows.

We find that CMU students are almost uniformly aware of the intellectual property issues dominating public discourse over the first half of 2012, as almost all respondents (93%) are aware of the SOPA/PIPA bills. In addition to their awareness, we wanted to measure the students' attitudes toward the bills. So, we provided a brief description of SOPA/PIPA for those who were not aware of them and then asked the respondents if they supported these policies or not. The answer options were on a standard 5-point Strongly Oppose to Strongly Support scale. The respondents who selected 'Oppose' or 'Strongly Oppose' constituted 82% of the sample, with 12% being neutral. A finding that could be of interest to government representatives voting on any future versions of these bills is that only 6% of our respondents support them, and only 1 person (<0.5%) expressed strong support. Interestingly, while such a large percentage oppose the bills, 68% are convinced that changes need to be made with respect to copyright protection policies. These percentages answer our question concerning students' views on piracy and whether or not the correct steps are being followed (or proposed) to prevent it.

Something we were expecting to observe is that students engage in piracy via streaming a lot more than by directly downloading content. Our results show that 77% of CMU undergraduates streamed pirated content (76% unweighted), whereas 75% have pirated by downloading (76% unweighted). The differences between streaming and downloading in our results are minimal. However, more detailed questions addressing these specific piracy behaviors should shed more light on the potential difference.

Only 30% of respondents were unaware when they were viewing/using/listening to pirated content. This shows that a majority of respondents are aware of their habits, but still do them regardless of potential consequences.

In terms of downloading pirated content, many people acknowledged that it is illegal (47%) and a fair amount of people agreed that it was unethical (34%). This contrasts with 51% of respondents who thought it was acceptable. Thus, it appears that over half of the respondents are actually OK with downloading pirated content.

However, when the question regarded simply the streaming of pirated content, only 27% believed that to be illegal, and a large portion now switched their responses thinking this was acceptable at 71%. Moreover, only 42% of respondents believe streaming pirated content to be illegal.

For each of these questions, roughly only a sixth of the respondents had no opinion on the subject.

(17% for downloading and 15% for streaming). These questions were meant to address people's attitudes toward different types of piracy and see if their opinions would change depending on the situation. Our results closely followed our hypothesis that people would be more accepting of streaming pirated content than downloading it.

We then proceeded to ask whether or not students' attitudes would change if the US was stricter about enforcing about pirated content. A large 44% said that their attitudes would not change at all. Of the remaining 56% that said their attitudes would in fact change, only 5% said his/her attitude would change a lot. 35% said their attitudes would change somewhat, which is what we expected. This question allows us to see what sort of reaction the students would have if the SOPA/PIPA bills would pass. It is fairly clear to us that the passing of the bills would either not change students' attitudes, or change them a little bit. Not many drastic measures would be taken.

Some interesting findings:

When looking at the responses, we decided to compare our results from similar questions to see how much consistency there was between questions that were trying to extract the same information. We compared the results from people's perceptions on downloading and streaming, and we compared those to the scenario based questions. First, comparing respondents' views on streaming and the youtube scenario question that focused on streaming, we found that 84.5% of all respondents who thought streaming was acceptable also thought that the youtube scenario on streaming was acceptable. This was fairly interesting even though the percentage was very high. We were interested that there was in fact a little bit of inconsistency for the responses to these two questions that should produce identical responses. Roughly 15% of the respondents were either inconsistent with their responses, or they thought that the two questions were in fact different. However, we do know there will definitely be variation from question to question. The percentage of people that thought these two questions regarding streaming were both illegal was 60.0%, which is must less than the previous proportion.

Now to look at respondents' perceptions on downloading, we compared the results from the two questions regarding their views on downloading pirated content and the scenario question about respondents' views on downloading a movie from dtella. We found that 91.0% of the people that though one was acceptable also thought the other was acceptable. This is a very large percentage and it is certainly a higher percentage than that for the streaming questions. It is sort of intriguing thought that still 9% of the people were inconsistent with their responses, as the dtella scenario question actually had the phrase "you download this movie from dtella." However, we do know that responses will undoubtedly vary from question to question and people taking the survey may forget previous responses. 75.9% of respondents agrees that both the question about downloading and the scenario were illegal.

Using these weights, we were able to find the post-stratification averages for different variables. The variable streaming was in response to the question, "Have you ever streamed?" where the respondents answered yes or no. We coded the variable to be one if the person had ever streamed

pirated content. An average weighted value of 0.07674 was found, which means that 77% of the population has streamed before (95% CI using jackknife variance: [0.7376845, 0.7911469]). The same approach was used for the variable which asked whether the respondent had downloaded or not. A weighted average value of 0.7535 was found meaning that around 75.3% of the population has downloaded pirated content (95% CI: [0.7253695, 0.7816727]). The variable policy support asked how the respondent felt about the proposed SOPA and PIPA bills on a scale of strongly oppose, oppose, neutral, support, strongly support. This levels were coded on a scale of 1-5 where 1 is strongly oppose and 5 is strongly support. A weighted average value of 1.807 was found, meaning that in general people are opposed to the proposed bills. The variables streaming frequency and downloading frequency ask how often the respondent streams or downloads on a scale of daily, weekly, few times a month, few times a year, and once. Those who answered 'no' to the question of have you ever streamed or downloaded were not asked this question, so those respondents were removed from the answer pool. For streaming frequency, there is an average frequency of 3.005 which means that in general students stream a few times a month. Similar results are found in the download frequency with a value of 2.697, so its also most common for someone to download a few times a month, given they download. Note that the average frequency for streaming is slightly higher than for downloading, suggesting that streaming is a more popular type of internet piracy on our campus.

Variable	$\bar{y}_w$	$Var(\bar{y}_w)$	$\bar{y}_{srs}$	$Var(\bar{y}_{srs})$	DEFF
Streaming	0.7674	0.0007378	0.7605	0.1821	0.0040
Downloading	0.7535	0.0008189	0.7595	0.1843	0.0044
Policy Support	1.807	0.003595	1.8109	0.7919	0.0045
Stream Frequency	3.005	0.006700	3.0388	1.0432	0.0064
Download Frequency	2.697	0.006303	2.7262	1.1100	0.0057

The barplot in the appendix visualizes the differences in downloading and streaming behaviors among class years. These are the questions that asked if a respondent has ever engaged in one of

these acts of piracy. Note that in the table above the weighted average for streaming is higher than the unweighted one, while for downloading it is the other way around. The weighted average is lower than the unweighted one. The barplot shows that more juniors have streamed, while more freshmen have downloaded. Since the post-stratification weights assign a weight of higher than 1 to juniors, it results in the aforementioned difference in the relationships of weighted and unweighted averages.

For quantitative data on downloaded content, the average response was “Few Times a year”, perhaps because it was the most convenient to choose (if respondents were not sure how many times they had downloaded pirated content).

Some interesting relationships:

We were curious as to how many people who thought streaming was acceptable (not acceptable) also thought downloading pirated content was acceptable (not acceptable), and to see how many respondents switched their answers between the two questions. We also wanted to test the significance of their responses. Therefore we created a fourfolds plot to test for this. The way you read a fourfolds plot is that there are four quadrants, each for the four combinations of the responses. The four quadrants are downloading and streaming are both acceptable (lower right), downloading and streaming are both not acceptable (upper left), downloading is acceptable but streaming is not acceptable (lower right), and downloading is not acceptable but streaming is acceptable (upper right). The opposite quadrants match up and create a 95% confidence interval (the black lines with upper and lower bounds), and if the confidence intervals do not overlap, the results are significant. Thus, we see that the two confidence intervals do not overlap and the results are significant. We can see here that much more people say that downloading and streaming are either both acceptable or both not acceptable. We see that between these two questions not many people changed their responses, which is what we originally hypothesized. We observe that 52 of the respondents thought streaming was acceptable, but when asked about downloading, switched their responses to not acceptable. Also, only 6 respondents thought streaming was not acceptable, but when asked about downloading, they actually switched their responses to acceptable. This is a very interesting outcome. 2.5% (6/240) believe that

downloading is worse than streaming.

We also wanted to take a look at whether males or females were more likely to have heard of the SOPA/PIPA bills, so we created a mosaic plot to take a look at these two variables. This mosaic plot shows two categorical variables, with the students' awareness of the bills as the marginal distribution, and gender as the conditional. If a box is shaded blue than there are significantly more values for that combination than expected, and if colored red, significantly fewer data points than expected. We see that the combination of the gender being female and the student not being aware of the bills is shaded blue, and thus has a significantly higher number of observations for that combination. We also see that the combination of males who have not heard of the bills has a significantly lower count. Therefore, it seems that females seem more likely to not have heard of the SOPA/PIPA bills as opposed to males.

We also performed some analysis on the variable political, which measures each respondent's political affiliation. We compared that with several different variables, some of which are: if you have heard of the bills, how much your support them, and whether or not you think changes need to be made with regard to piracy. We constructed several graphical displays, such as a stacked bar chart and a mosaic plot, however no association between any of the variables and someone's political affiliation could be found. The results seemed to be pretty standard and even across the board.

In order to investigate a possible relationship between ethnicity and the other variables, and initially to decide whether we had to weight by ethnicity, we produced several mosaic plots. Ethnicity was a variable on the survey in which you were able to select any of the ethnicities that applied to you from a list of African American, Asian/Pacific Islander, Caucasian, Hispanic/Latino, or Native American. Because we allowed the respondent to pick as many of the ethnicities as they pleased, it created a large amount of levels for the ethnicity variable, many of which had a count of less than five. In order to analyze this variable we changed the respondents' ethnicity into only one ethnicity. Due to the fact that we didn't know which of the checked ethnicities was the respondent's dominant ethnicity, we duplicate the person's result, giving them both ethnicities. We did this with conditionality to three different variables, bill awareness,

streaming, and downloaded. When looking at the three mosaic plots we see that there are no significant residuals, so we would conclude that there is no significant relationship between ethnicity and bill awareness, streamed, and downloaded. This means that we don't need to weight by ethnicity.

We took the same approach in finding if we should weight by school, which it ended up that school was insignificant as well in the mosaic plots, and we did not need to weight by school.

Section 4: Discussion

Comments from participants were diverse. Some respondents expressed opposition to piracy while others touted their abstinence from the practice. There were also those who had pirated in the past, but stopped after copyright infringement incidents, e.g., getting blocked from the CMU internet. In general, the open responses reflected our quantifiable data.

Respondents stated that limited offerings from on-demand subscriptions like Netflix and Hulu precipitated their pirating. The internet's expedience proved to be an incentive to download/stream. For example, Netflix does not offer "just-aired" episodes; it only offers episodes from past seasons. Naturally, the cost was a factor in pirating: pirating is free, while a Hulu subscription, for example, is not. Regardless of individual behaviors, some respondents agreed that the quality/proliferation of media will degrade if piracy continues in the long run.

Some respondents remained optimistic about media consumption: a couple of comments indicated in one way or another that if someone is passionate about the artistic integrity of a piece of music or a film, they will purchase it legally out of principle: "Sure, internet piracy may take profits away from big businesses/artists, but if people really like a product enough, they will be willing to buy it." Others suggested that as long as downloading continues, artists will obtain greater fame and exposure than they would have otherwise: "The industry and copyright laws need to adapt to EMBRACE the internet era. All of these things that are considered illegal can serve to help the industry if they do it right." Moreover, some people cited that media from non-mainstream (i.e indie films) sources do succeed immensely from internet-based exposure.

Respondents also stated that streaming media helps decide whether or not one wants to purchase product (pre-evaluation of the quality of product).

Overall, the open responses showed that respondents were aware of and opinionated about piracy and the SOPA/PIPA bills. Participants felt more vulnerable than guilty about the recently proposed bills. Others felt that the SOPA/PIPA bills would hinder the creativity of artists/producers based on the internet. Many decided not to take responsibility for their actions and stated that the piracy policies should blame internet service providers (ISPs) rather than individual users. It should be noted that the proposed legislation would have bestowed ISPs with a greater burden to police the internet. Some respondents qualified or defended their own behaviors and claimed that simply viewing (streaming) pirated material is not as bad as actually disseminating or even downloading it.

Our data supports the notion that undergraduates believe downloading pirated content is more illegal, less ethical, and less acceptable than streaming pirated content. Thus, undergraduates may believe that viewing "shared" media (material attached to a media server) is less offensive than downloading copyrighted material to one's own hard drive.

The majority of undergraduates stated that copyright regulations need to change in the context of the internet. However, most respondents indicated that *if* the United States were to change piracy laws, undergraduate attitudes about piracy would *not* change. These beliefs--that (a) regulations need to be changed and (b) views about piracy would not change regardless of regulations--are somewhat contradictory; however, it could suggest that laws do not influence opinions and behaviors regarding the internet.

Our survey indicates that students embrace instant access to material on the internet: "You know, the internet age has sort of ushered in this attitude that we deserve everything we want for free, and we want it instantly." Piracy and torrents have become an illegitimate alternative to traditional sources of entertainment like television and movie theaters. However, the age of the internet has not signaled the end of legitimate entertainment, but rather, ushered in a wave of on-demand streaming-based media like Netflix and Hulu: Netflix earnings exceeded $1.5 billion in 2011 (Sandoval). Subscriptions to these services provide convenient access to a wide range of media for a relatively low price.

It is not exactly unexpected that people would feel more comfortable pirating from a site like YouTube (a Google subsidiary) rather than the back-alleys of the internet. However, it is

interesting to ask "why", given the outcome of the survey so far. While you can argue that YouTube is built on a foundation of videos that have violated copyrights in one way or another, in its early days and still today, many of the videos are created by its users. The original content on the site provides the website with some legitimacy, relative to sites that have been successfully sued like Napster. YouTube developed itself into something a "brand" and a "destination site"(Flores); this could be why Google Video never did as well as YouTube: Google acquired YouTube for $1.65B, 10/2006. With newfound deep pockets, YouTube found itself the center of attention for potential litigation. Suing a (for the most part) user-operated site remains complicated, but YouTube has had to negotiate with companies like Sony, Universal, CBS, and the NHL to give them what amounts to free advertising (Breen). While Viacom had its law suit dismissed against Google, an American appeals court recently found sufficient evidence of YouTube's awareness of its facilitation of copyright infringement; they are also facing legal troubles over the same issue in Germany (Schwartz). There is no doubt that the online-copyright climate is changing, but, while their legal issues remain unresolved, it does not seem like YouTube is going away as long as it does a good job of placating detractors. So while much of the content on the site is in a legal grey area, there it retains some legitimacy, and this is very clear in our findings: 82% of respondents indicated it was acceptable to watch what amounts to pirated material on YouTube, while just 61% of respondents considered streaming pirated content on the internet-in-general acceptable.

It appears that our research questions garnered the information that we hoped to solicit from respondents. A majority of respondents choosing to express themselves in the optional comments portion of the survey shows that our survey's content is relevant to our target population. However, this indicates the possible presence of a strong self-selection bias among those who accepted our survey solicitation.

In terms of developing our survey, we found distributing early drafts to members of our target population for informal feedback helpful: early feedback indicated that we should separate "streaming" from "downloading". The differences in respondent perception of downloading and streaming turned out to be one of the most striking aspects of the survey. Furthermore, we were able to generate a 66% response rate while running the pre-test at Hunt Library (the response rate could have been considered 100%, depending on measurement technique; some people we approached passed the survey to a friend, but the survey was still completed). We had some

problems comparing our ethnicity categories (we had 6) to the CMU Factbook ethnicity categories (4). Furthermore, because a number of individuals reported multiple ethnicities, it was hard to compare our demographic information

The main conclusion we can make thus far is that the majority of the respondents strongly oppose the SOPA/PIPA bills, but feel that changes need to be made with respect to copyright protection on the internet. Regarding one of our initial questions, whether or not students oppose the bill because they like to pirate or because they are concerned about the integrity of the internet in the hands of government, these quotes help sum up student opinions:

> *"The US government can barely keep their budget intact, I do not feel giving them authority over the internet will be done in the correct manner"*

> *"Take down illegal content but don't mess with the open infrastructure of the internet."*

It appears that almost all CMU students engage in piracy, most are not unabashed about doing so, and there are some who are genuinely concerned about the democracy of the internet.

Appendix

Appendix

Equations:

Weighted Averages:

$$\bar{y}_w = \frac{\sum_i w_i y_i}{\sum_i w_i} \quad (1)$$

$$\bar{y}_w^{(r)} = \frac{\sum_{i=1}^{n} w_i^{(r)} y_i^{(r)}}{\sum_{i=1}^{n} w_i^{(r)}} \quad (2)$$

$$\bar{y}_{JK} = \frac{1}{n}\sum_{r=1}^{n} \bar{y}_w^{(r)} \quad (3)$$

Variances:

$$\mathrm{Var_{JK}}(\bar{y}_w) = \frac{\mathrm{n}-1}{\mathrm{n}}\sum_{\mathrm{r=1}}^{\mathrm{n}} \left(\bar{y}_w^{(r)} - \bar{y}_{JK}\right)^2 \quad (4)$$

Confidence Intervals:

$$CI: \left(\bar{y}_w - 1.96\sqrt{\left(1-\frac{\mathrm{n}}{\mathrm{N}}\right)\mathrm{Var}(\bar{\mathrm{y}}_\mathrm{w})}, \bar{y}_w + 1.96\sqrt{\left(1-\frac{\mathrm{n}}{\mathrm{N}}\right)\mathrm{Var}(\bar{\mathrm{y}}_\mathrm{w})}\right) \quad (5)$$

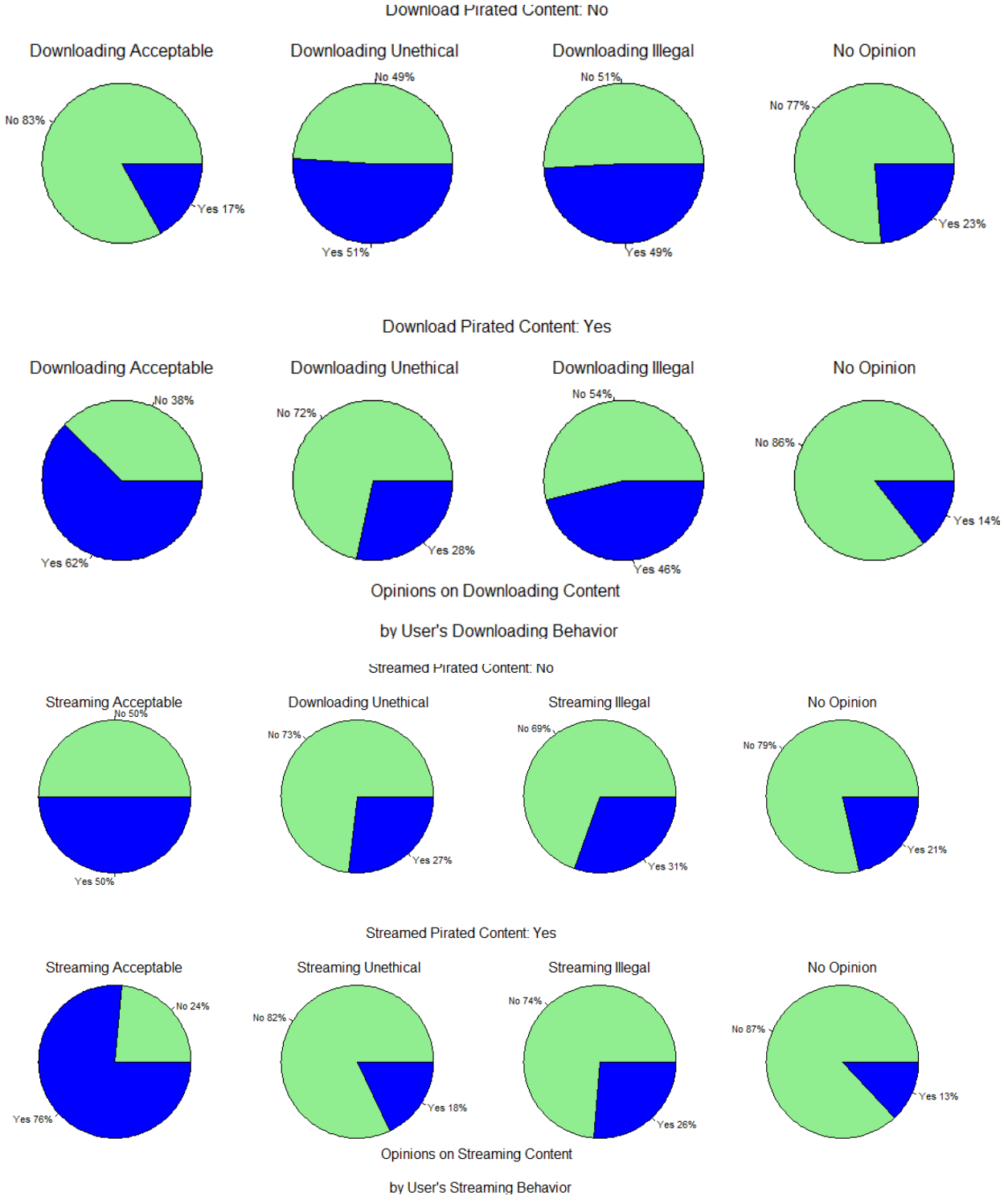
Sample Size:

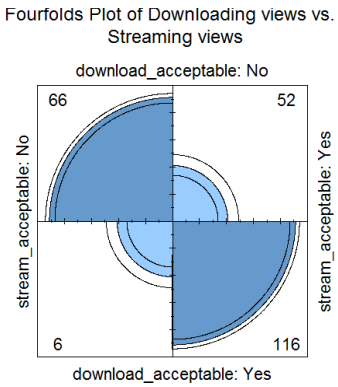
$$n = \frac{Nn_0}{N+n_0} \rightarrow n_0 = \frac{Nn}{N-n}$$

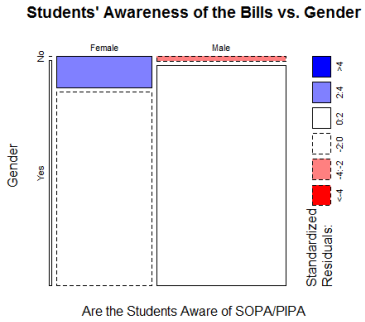
Margin of Error:

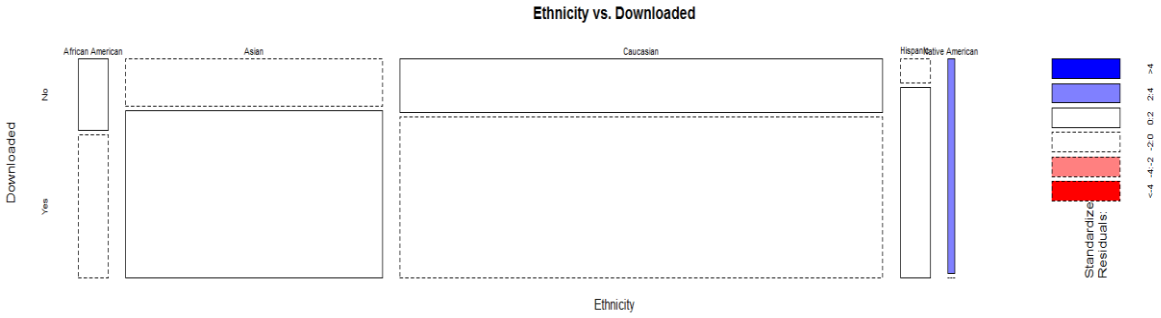
$$MOE = z_{0.025}\sqrt{\frac{p(1-p)}{n_0}}$$

Graphs

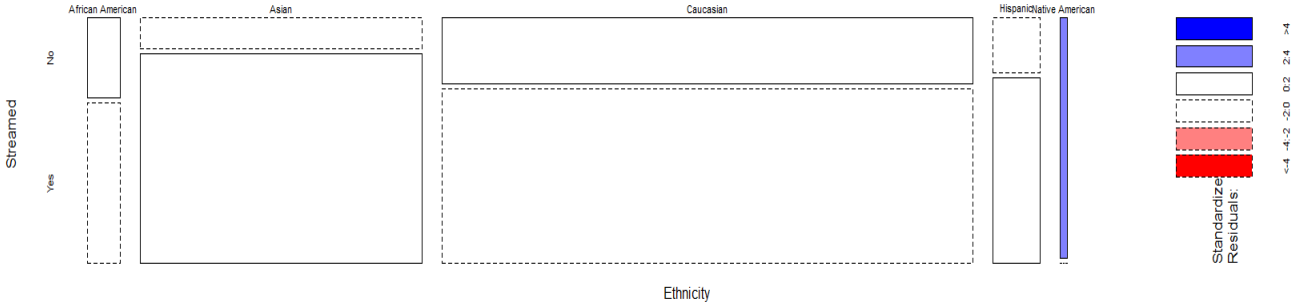

Download Pirated Content: No
Downloading Acceptable
No 83%
Yes 17%
Downloading Unethical
No 49%
Yes 51%
Downloading Illegal
No 51%
Yes 49%
No Opinion
No 77%
Yes 23%
Download Pirated Content: Yes
Downloading Acceptable
No 38%
Yes 62%
Downloading Unethical
No 72%
Yes 28%
Downloading Illegal
No 54%
Yes 46%
No Opinion
No 86%
Yes 14%
Opinions on Downloading Content
by User's Downloading Behavior
Streamed Pirated Content: No
Streaming Acceptable
No 50%
Yes 50%
Downloading Unethical
No 73%
Yes 27%
Streaming Illegal
No 69%
Yes 31%
No Opinion
No 79%
Yes 21%
Streamed Pirated Content: Yes
Streaming Acceptable
No 24%
Yes 76%
Streaming Unethical
No 82%
Yes 18%
Streaming Illegal
No 74%
Yes 26%
No Opinion
No 87%
Yes 13%
Opinions on Streaming Content
by User's Streaming Behavior


Fourfolds Plot of Downloading views vs. Streaming views
download_acceptable: No
66
52
6
116
stream_acceptable: No
stream_acceptable: Yes
download_acceptable: Yes

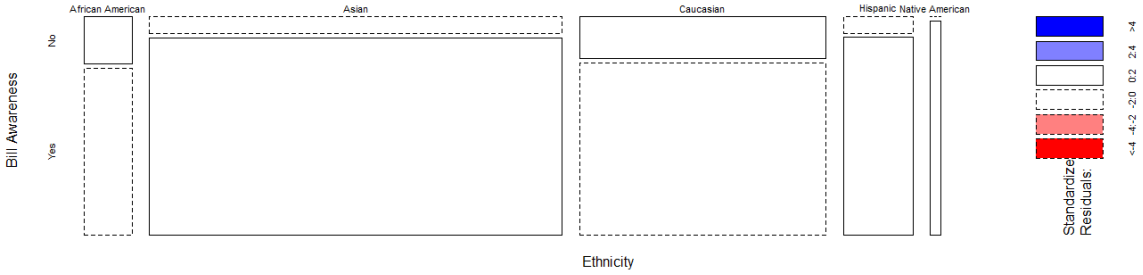

Students' Awareness of the Bills vs. Gender
Female
Male
No
Yes
Gender
Standardized Residuals:
>4
2:4
0:2
-2:0
-4:-2
<-4
Are the Students Aware of SOPA/PIPA


Ethnicity vs. Downloaded
African American
Asian
Caucasian
Hispanic
Native American
No
Yes
Downloaded
Standardized Residuals:
>4
2:4
0:2
-2:0
-4:-2
<-4
Ethnicity

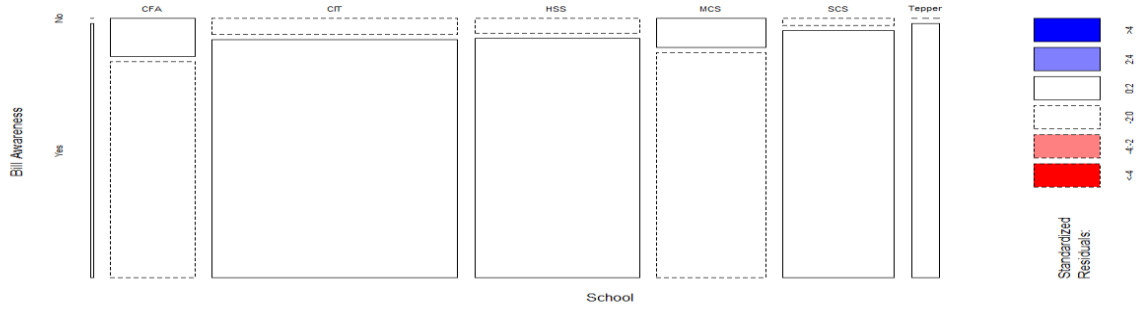
Ethnicity vs. Streaming



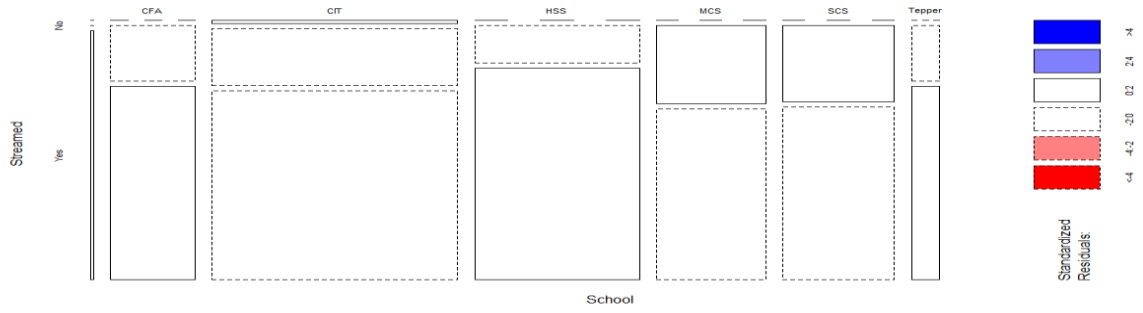
Ethnicity vs. Bill Awareness

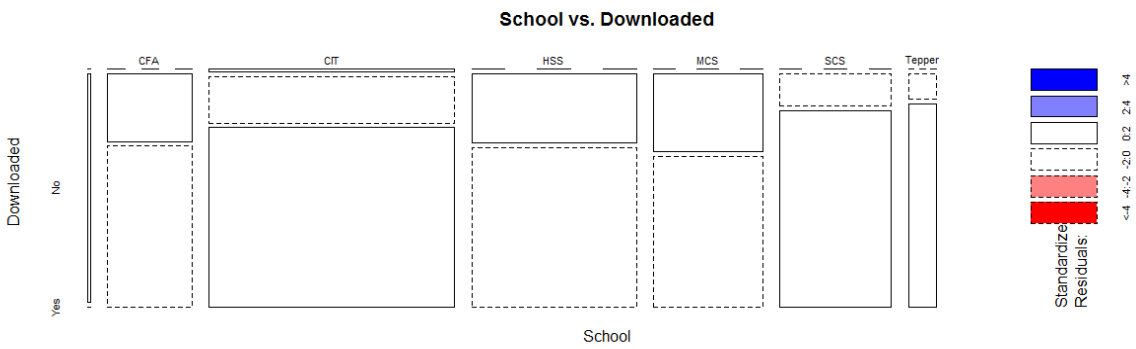


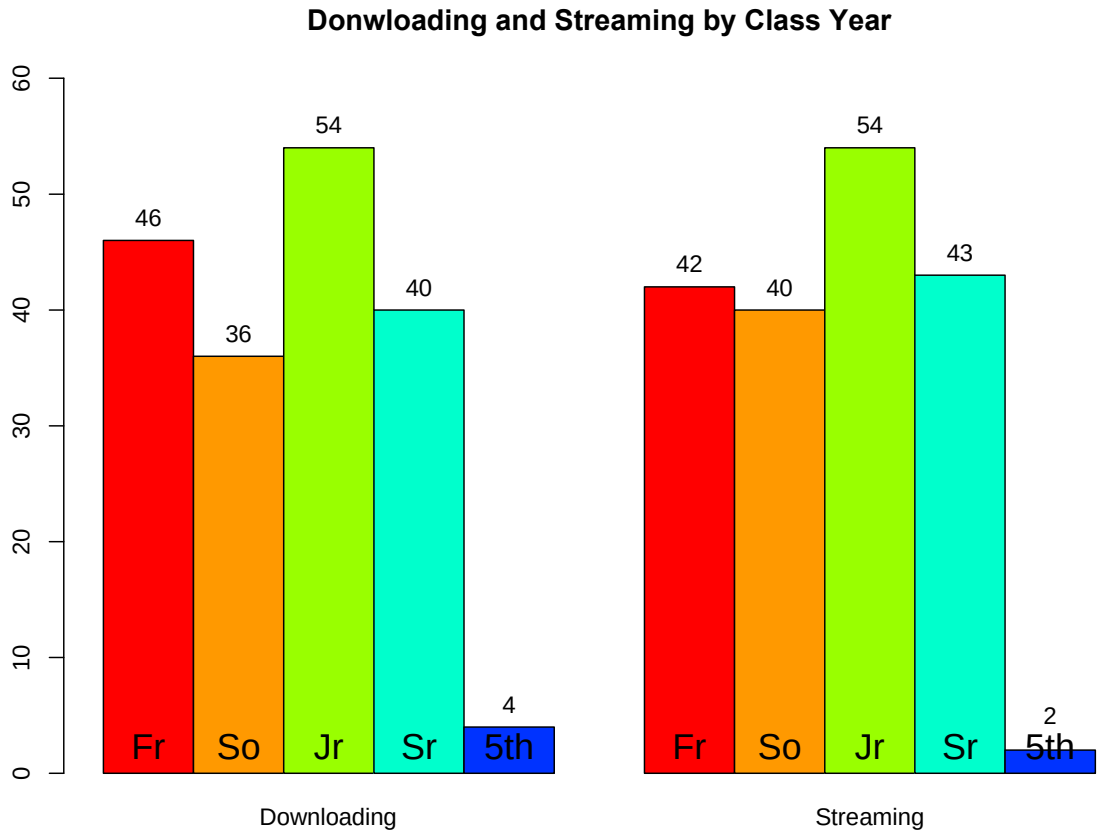
School vs. Bill Awareness



School vs. Streamed




School vs. Downloaded
CFA
CIT
HSS
MCS
SCS
Tepper
Downloaded
No
Yes
School
Standardized Residuals
<-4
-4:-2
-2:0
0:2
2:4
>4


Donwloading and Streaming by Class Year
0
10
20
30
40
50
60
46
36
54
40
4
42
40
54
43
2
Fr
So
Jr
Sr
5th
Fr
So
Jr
Sr
5th
Downloading
Streaming

Survey

CMU Internet Piracy Survey

In recent months, internet piracy and copyright enforcement have been at the forefront of legislative debates, media coverage, and corporate public statements. We are surveying Carnegie Mellon undergraduates to see how often they engage in internet piracy and what views they hold on related copyright enforcement legislation. You will be able to enter your Andrew ID on the last page of the survey if you want to participate in the raffle for a chance to win an Amazon gift card.
* Required

Have you heard of the recently proposed SOPA/PIPA bills? *
- Yes
- No

The SOPA Bill involves the following: - Expansion of the ability of U.S. law enforcement to fight online trafficking in copyrighted intellectual property and counterfeit goods - Expansion of existing criminal laws to include unauthorized streaming of copyrighted content, with a maximum penalty of a five-year prison sentence - Allowance for court orders to bar advertising networks and payment facilities from conducting business with infringing websites, block search engines from linking to the offending sites, and require Internet service providers to block access to the sites The PIPA Bill involves: - Prevention of real online threats to economic creativity and theft of intellectual property - Provisions for the U.S. government and copyright holders to utilize additional tools to curb access to online distribution and sales of pirated content and counterfeit goods, especially those registered outside the U.S.

Do you support the policies outlined in the definitions of the proposed bills?
- Strongly Oppose
- Oppose
- Neutral
- Support
- Strongly Support

Do you feel changes need to be made with respect to copyright protection in the internet era?
- Yes
- No

Have you ever downloaded pirated content?‘Pirated content’ being copyrighted material used/reproduced without the consent of the copyright holder. Example: Download a movie from a torrent.
- Yes
- No

How often do you download pirated content?
- Daily
- Weekly
- Few times a month
- Few times a year
- Only once

Have you ever streamed pirated content?Example: Stream a TV show from Megavideo/SideReel.
- Yes
- No

How often do you stream pirated content?
- Daily
- Weekly
- Few times a month

- Few times a year
- Only once

Do you use ... ?Check all that apply.

- Streaming sites (e.g. sidereel.com)
- Peer-to-peer file-sharing client(s) (e.g. Dtella, LimeWire)
- Direct download sites (e.g. rapidshare.com)
- I do not use any of the above

Do you usually know whether the content you are viewing/using/listening to is being reproduced illegally or not?

- Yes
- No

Do you consider downloading pirated content to be...?Check all that apply.

- Acceptable
- Illegal
- Unethical
- No opinion on the ethics or legality of this action

Do you consider streaming/viewing/listening to unlicensed content (without downloading) to be...?Check all that apply.

- Acceptable
- Illegal
- Unethical
- No opinion on the ethics or legality of this action

For example, do you consider viewing/listening to music on YouTube that has NOT been posted by the original artist (i.e. NOT through a VEVO or official artist account) to be...?Check all that apply.

- Acceptable
- Illegal
- Unethical
- No opinion on the ethics or legality of this action

If the US adopted a tougher stance on internet piracy enforcement, to what degree would your attitudes about using pirated content change?

- Not at all
- Somewhat
- A lot
- Unsure

Scenarios

Suppose you want to watch a movie that you have anticipated for a long time.

You download this movie from dtella (or another peer-to-peer file-sharing network) and watch it on your computer. Do you consider this to be...?Check all that apply.

- Acceptable
- Illegal
- Unethical
- No opinion on the ethics or legality of this action

You cannot find the movie on dtella, but you notice it has been uploaded in multiple parts to YouTube. You watch the movie in its entirety on YouTube. Do you consider this to be...?Check all that apply.
- Acceptable
- Illegal
- Unethical
- No opinion on the ethics or legality of this action

After watching the movie on YouTube, you save it to a file on your computer and upload the file to a sharing website. You personally do not profit in any way from this activity. Do you consider this to be...?Check all that apply.
- Acceptable
- Illegal
- Unethical
- No opinion on the ethics or legality of this action

Do you think there is a way to watch the movie that is as convenient as the methods above but requires payment (potential sources include iTunes, Netflix, etc.)?
- Yes
- No

Do you have any other comments or views on internet piracy that this survey has not yet addressed? Please express them here.

Have you had any personal experiences that caused a change in your attitude towards internet piracy and copyright enforcement? Please describe them here.

Demographic Information

Gender
- Male
- Female

What is your ethnicity?Check all that apply.
- White/Caucasian
- African American
- Hispanic/Latino
- Asian/Pacific Islander
- Native American
- Other:

What is your class year?
- Freshman
- Sophomore
- Junior
- Senior
- 5th-year senior

Do you live on campus or off campus?
- On campus

- Off campus

What is your political affiliation?
- Republican
- Democrat
- Independent
- Other
- None
- Do not wish to share to information

What school are you in?(If you are in multiple schools, please select the school of your primary major.)
- CFA
- HSS
- CIT
- SCS
- Tepper
- MCS

If you are from the United States, or have lived in the United States for the majority of your life, what region are you from?If you are not from the US, please pick Other and write down where you consider yourself from.
- Northeast
- Midwest
- South
- West
- Other:

To participate in the raffle enter your Andrew ID in the box below.We assure you that your ID will not be linked to your responses

Email

Dear CMU Undergraduate,

The students of 36-303 (Sampling, Surveys, & Society) are conducting a short, confidential survey regarding students' opinions on internet piracy and copyright enforcement. You've been randomly selected from all CMU undergraduates to participate and we would really appreciate your help.

Take the survey here:
https://docs.google.com/spreadsheet/viewform?formkey=dExYMFBneHNvY0hzOWh1VW5hY0Q0VWc6MQ

Two participants will be chosen at random to receive $30 Amazon gift cards. In order to be eligible for the raffle, you will be required to enter your Andrew ID upon completion of the survey; if you do not wish to be eligible, you may refrain.

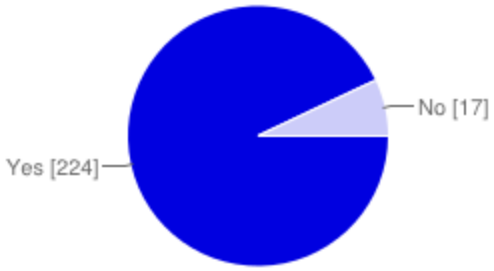
Regardless of raffle participation, your responses are voluntary and will be confidential. Responses will not be identified by individual. The survey will take about three minutes to complete, and you can exit the survey or choose to skip any questions you like.

Thank you for your participation!

Sincerely,
The Students of 36-303 [Group B]

Summary See complete responses

Have you heard of the recently proposed SOPA/PIPA bills?



Yes	224	93%
No	17	7%

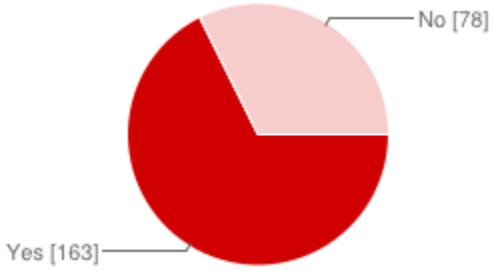
The SOPA Bill involves the following: - Expansion of the ability of U.S. law enforcement to fight online trafficking in copyrighted intellectual property and counterfeit goods - Expansion of existing criminal laws to include unauthorized streaming of copyrighted content, with a maximum penalty of a five-year prison sentence - Allowance for court orders to bar advertising networks and payment facilities from conducting business with infringing websites, block search engines from linking to the offending sites, and require Internet service providers to block access to the sites The PIPA Bill involves: - Prevention of real online threats to economic creativity and theft of intellectual property - Provisions for the U.S.

government and copyright holders to utilize additional tools to curb access to online distribution and sales of pirated content and counterfeit goods, especially those registered outside the U.S.

Do you support the policies outlined in the definitions of the proposed bills?

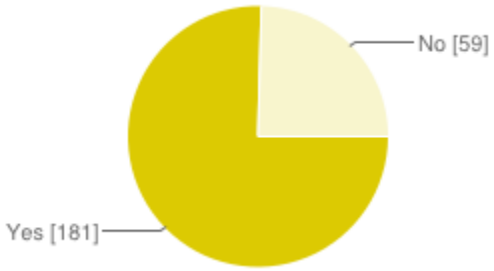
Strongly Oppose	105	44%
Oppose	91	38%
Neutral	30	12%
Support	14	6%
Strongly Support	1	0%

Do you feel changes need to be made with respect to copyright protection in the internet era?



Yes	163	68%
No	78	32%

Have you ever downloaded pirated content?



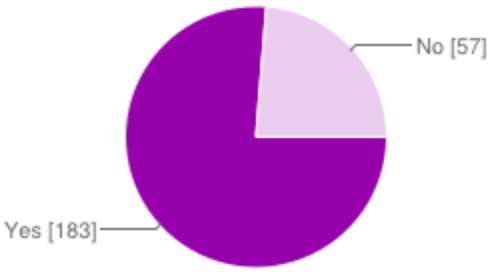
Yes	181	75%
No	59	24%

How often do you download pirated content?

Daily	13	5%

Weekly	29	12%
Few times a month	46	19%
Few times a year	78	32%
Only once	14	6%

Have you ever streamed pirated content?



Yes	183	76%
No	57	24%

How often do you stream pirated content?

Daily	14	6%
Weekly	46	19%
Few times a month	62	26%
Few times a year	51	21%
Only once	9	4%

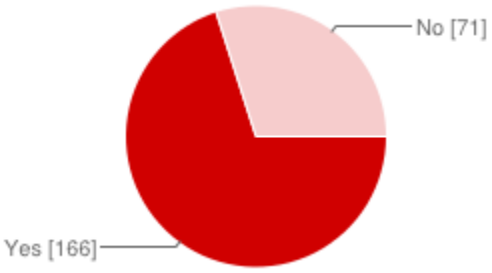
Do you use ... ?

Streaming sites (e.g. sidereel.com)	163	68%
Peer-to-peer file-sharing client(s) (e.g. Dtella, LimeWire)	117	49%
Direct download sites (e.g. rapidshare.com)	93	39%

I do not use any of the above	36	15%

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

Do you usually know whether the content you are viewing/using/listening to is being reproduced illegally or not?



Yes	166	69%
No	71	29%

Do you consider downloading pirated content to be...?

Acceptable	122	51%
Illegal	113	47%
Unethical	82	34%
No opinion on the ethics or legality of this action	40	17%

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

Do you consider streaming/viewing/listening to unlicensed content (without downloading) to be...?

Acceptable	168	70%
Illegal	65	27%
Unethical	49	21%
No opinion on the ethics or legality of this action	36	15%

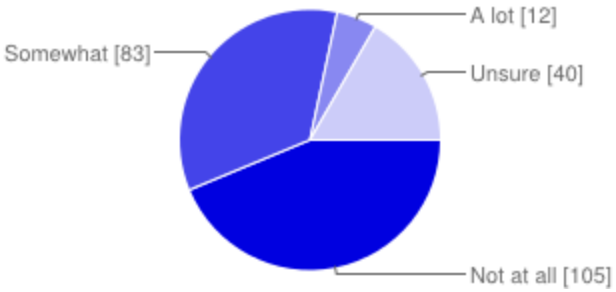
People may select more than one checkbox, so percentages may add up to more than 100%.

For example, do you consider viewing/listening to music on YouTube that has NOT been posted by the original artist (i.e. NOT through a VEVO or official artist account) to be...?

Acceptable	195	81%
Illegal	29	12%
Unethical	21	9%
No opinion on the ethics or legality of this action	34	14%

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

If the US adopted a tougher stance on internet piracy enforcement, to what degree would your attitudes about using pirated content change?



Not at all	105	44%
Somewhat	83	34%
A lot	12	5%
Unsure	40	17%

Scenarios

Suppose you want to watch a movie that you have anticipated for a long time.

You download this movie from dtella (or another peer-to-peer file-sharing network) and watch it on your computer. Do you consider this to be...?

Acceptable	140	59%
Illegal	100	42%
Unethical	75	31%
No opinion on the ethics or legality of this action	28	12%

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

You cannot find the movie on dtella, but you notice it has been uploaded in multiple parts to YouTube.

You watch the movie in its entirety on YouTube. Do you consider this to be...?

Acceptable	176	74%
Illegal	59	25%
Unethical	52	22%
No opinion on the ethics or legality of this action	18	8%

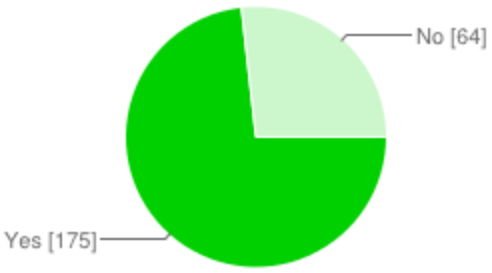
People may select more than one checkbox, so percentages may add up to more than 100%.

After watching the movie on YouTube, you save it to a file on your computer and upload the file to a sharing website. You personally do not profit in any way from this activity. Do you consider this to be...?

Acceptable	78	33%
Illegal	118	50%
Unethical	89	38%
No opinion on the ethics or legality of this action	38	16%

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

Do you think there is a way to watch the movie that is as convenient as the methods above but requires payment (potential sources include iTunes, Netflix, etc.)?



Yes	175	73%
No	64	27%

Do you have any other comments or views on internet piracy that this survey has not yet addressed? Please express them here.

Some movies I can easily find on iTunes and download/rent there, but I often watch movies that have yet to be released via illegal sources; Not really.There is obvious room for improvement in our legislature on piracy, however the bills proposed have some sections that are just absurd. Overall though, it seems like we need something to get people to stop easily stealing anything they ever

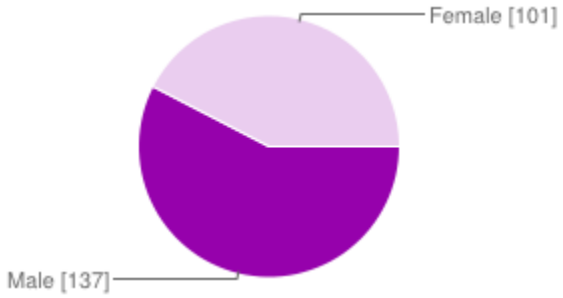
want.Without payment, the movie and music industry would not exist. So, if all people pirated this material online, the industries would not make a profit and would stop production. Thus, without some sort ...

Have you had any personal experiences that caused a change in your attitude towards internet piracy and copyright enforcement? Please describe them here.

Being part of the online tech community does foster a sense of openness and free media."Steam" (the largest digital distribution platform for video games) gained popularity some years ago, and—at least for me—made it easier to legally+ethically buy games than to steal them. I haven't stolen a game since I starting using Steam. Of course, I'm also no longer twelve, so that's probably a factor too. Just reading reddit and seeing how dumb people are trying to convince themselves that stealing content from big stars isn't bad.I come from a media family, so the idea of original and copyrighted w...

Demographic Information

Gender



Male	137	57%
Female	101	42%

What is your ethnicity?

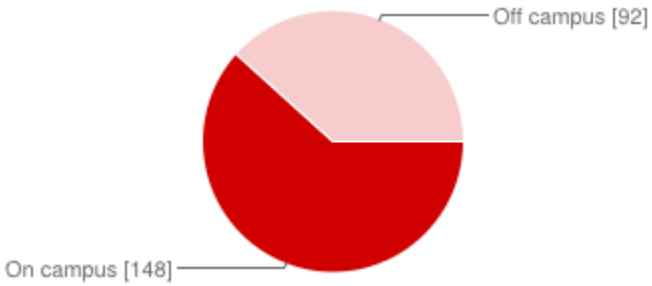
White/Caucasian	147	63%
African American	9	4%
Hispanic/Latino	13	6%
Asian/Pacific Islander	78	33%
Native American	3	1%
Other	2	1%

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

What is your class year?

Freshman	66	27%
Sophomore	55	23%
Junior	47	20%
Senior	64	27%
5th-year senior	7	3%

Do you live on campus or off campus?



On campus	148	61%
Off campus	92	38%

What is your political affiliation?

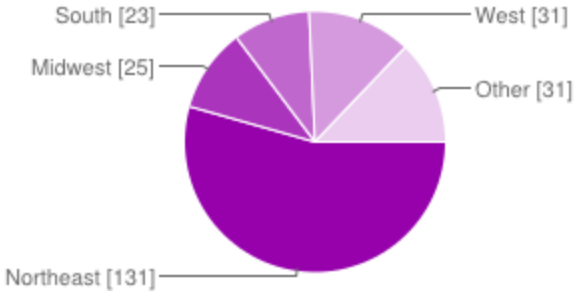
Republican	22	9%
Democrat	81	34%
Independent	35	15%
Other	17	7%
None	73	30%
Do not wish to share to information	13	5%

What school are you in?

CFA	28	12%

HSS	53	22%
CIT	79	33%
SCS	36	15%
Tepper	9	4%
MCS	35	15%

If you are from the United States, or have lived in the United States for the majority of your life, what region are you from?

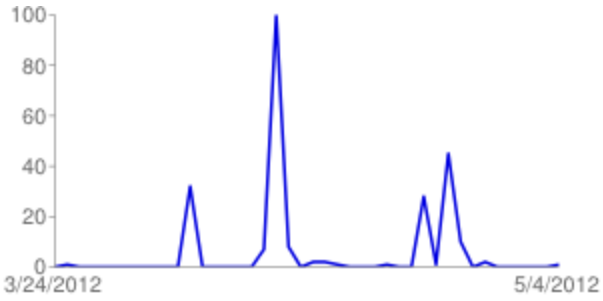


Northeast	131	54%
Midwest	25	10%
South	23	10%
West	31	13%
Other	31	13%

To participate in the raffle enter your Andrew ID in the box below.

zgallupanshulgorgouldenbrgormandswensgrahamt sgotschecmsundayngorskimgallabrptglbyronrlbrownejgoodwyntswansonjburrowsnsundaremburnettbg oodkswartsjordang1ncbmswishermhgrayaeburtonegonzalekbudwalrandy...

Number of daily responses



Bibliography

Edward Wyatt. “Lines Drawn on Antipiracy Bills”. *The New York Times.* December 14, 2011. See: http://www.nytimes.com/2011/12/15/technology/lines-are-drawn-on-legislation-against-internet-piracy.html?pagewanted=all

Niko Flores, How YouTube Works (Oct. 11, 2006). See: http://ondemandmedia.typepad.com/odm/2006/1 0/how-youtube_wor.html
Jason Breen, YouTube or YouLose: Can YouTube Survive a Copy Infringement Lawsuit? (2007) See: http://igeneration.edublogs.org/files/2008/09/youtube-or-youlose.pdf

Schwartz, Jan. “YouTube Dealt Fresh Blow by German Copyright Ruling.” *Reuters*. April 20, 2012. http://www.reuters.com/article/2012/04/20/net-us-youtube-germany-idUSBRE83J0NP20120420
Greg Sandoval. “Netflix Earnings Today: It’s Nail Biting Time.” *CNET News*. April 24, 2012. See: http://news.cnet.com/8301-1023_3-57418857-93/netflix-earnings-today-its-nail-biting-time/