Gaming Patterns in Higher Education

Author 1 and Author 2

2024-10-25

Header

Author Contributions

Author 1: Did the data cleaning and preliminary analysis, as well as contributed to questions #1 and #2, along with doing the advanced data analysis.

Author 2: Contributed to additional data cleaning specific to the questions and analysis for questions 3, 4, and 5.

Use of GPT

ChatGPT was used as a substitute for documentation for R. Since we were unfamiliar with R, we asked ChatGPT how to use R in certain methods in order to find and filter out conditions in the dataset. We additionally used GPT to analyze reasoning and to confirm what we thought was correct about the dataset, as well as to identify extra questions that could be answered for our advanced analysis.

Introduction

The data provided is from a survey done at UC Berkeley for students enrolled in a statistics course in 1994. Out of an eligible 314 students, 91 students submitted eligible responses which will be the data used in this EDA. The data focuses on student's play time of video games, likes and dislikes of video games, and more. One important thing to note is that responses that were left unanswered or improperly answered were coded as 99, and students who had never played video games or expressed a strong dislike for them were instructed to skip many questions. Additionally, the survey was distributed into two parts, where the first part asks about variable data, while the second part focuses on attitude questions.

Main Research Questions

1. Provide an estimate for the fraction of students who played a video game in the week prior to the survey.

2. Check to see how the amount of time spent playing video games in the week prior to the survey compares to the reported frequency of play (daily, weekly, etc). Additionally, consider the fact that there was an exam the week prior to the survey.

3. Provide a point estimate and an interval estimate for the average amount of time spent playing video games in the week prior to the survey while keeping in mind the overall shape of the sample distribution.

4. Consider the attitude questions. In general, do you think the students enjoy playing video games? If you had to make a short list of the most important reasons why students like/dislike video games, what would you put on the list?

5. Look for the differences between those who like to play video games and those who don't. To do this, use the questions in the last part of the survey, and make comparisons between different students and their categories.

6. Further investigate the grade that students expect in the course. How does it match the target distribution used in grade assignment of 20% A's, 30% B's, 40% C's and 10% D's or lower? If the nonrespondents were failing students who no longer bothered to come to the discussion section, would this change the picture?

Basic Analysis

Data Processing and Summaries

In order to analyze our data, we wanted to make sure that our data was clean, and was able to have analysis performed on it. In order to do this, we looked through the data to identify if there were any inconsistencies that we wanted to catch and either fix or get rid of. One such inconsistency was included when we went to check the data types for the second part of the survey.

Through this, we noticed that while the first part of the survey all contained numeric types, which was to be expected, the second part of the survey contained two non-numeric columns, 'other' and 'other2', both of which were not supposed to be included in the second part of the survey. In order to clean the data up, we removed these two columns.

```
survey_p2 <- survey_p2[, !(names(survey_p2) %in% c("other", "other2"))]
all(sapply(survey_p2, is.numeric))
```

```
## [1] TRUE
```

Furthermore, in order to clean up the data and keep it consistent, we wanted to combine our data into one data frame. By doing this, we are using the assumption that each row in 'survey_p1' is the same person that was in the same row in 'survey_p2', meaning that they are the same person. Additionally, we also got rid of rows that contained NA values, just to make sure we got rid of inconsistent data, because that could mess up our survey. From this, we noticed that we have 87 rows of data left, which means that 4 of our rows had NA values.

```
combined_table <- cbind(survey_p1, survey_p2)
cleaned_df <- na.omit(combined_table)
dim(cleaned_df)
```

```
## [1] 87 34
```

We also noted down the types of variables that were present in our data, since many were binary, ordinal, nominal, and more. We also graphed out histograms of the variables in the two surveys to see if we could identify any outliers, but since responses that were left unanswered or were improperly answered were coded as 99, there were no such outliers. Therefore, we went with the assumption that there were no outliers in the dataset, since they were all coded as 99 and could be filtered out later.

Question 1: Estimate of Students who played a Video Game

Methods

In order to provide a solid estimation of the amount of people who played a video game in the week prior to the survey, let's filter the amount of people who had over 0 hours last week as well as filter out those who didn't respond correctly or did not answer the question (99)

```
## [1] "Point Estimate:  0.3678"
```

Furthermore, in order to get a more concrete range for the true measurement of this estimate, let's take a 95% confidence interval and get an interval estimate for the true population proportion who played a video game during an exam week.

```
## [1] "95% Confidence Interval: [0.2665, 0.4691]"
```

Analysis

Our examination of video game engagement among UC Berkeley Statistics students during the Fall 1994 semester revealed that 36.78% of surveyed students reported playing video games in the week following their second examination. The 95% confidence interval for this proportion ranges from 26.65% to 46.91%, indicating the true population parameter lies within this range with 95% confidence.

Key analytical findings include:

- The point estimate of 36.78% represents a substantial portion of the student population engaging in gaming activities in an academically intense period.
- The confidence interval width of ~20% suggests a high variability amongst students in terms of gaming behavior.
- The timing of the survey, during an exam week, provides context for interpreting these gaming patterns.

Conclusions

Based on our analysis, we can draw several significant conclusions:

- Video gaming was a common activity amongst Statistics students, even in a academically intense time. Our confidence interval shows us that at minimum, 25% of the student population was engaged in gaming activities during the survey period.
- Video games were played by a significant amount of people in the Statistics course, yet our wide confidence interval suggests that there is a lot of variability amongst the people in the course, possibly suggesting that difference populations of students would have a different estimate about the percentage of students who played video games regularly.
- The percentage found from the data indicates that gaming was a prevalent hobby even in the mid-1990s, and can perhaps provide insight into student gaming behavior into the future, as perhaps it has become even more significant as time has passed and gaming has become easier to access for everyone.

Question 2: Find Effect of Exam Week on Time Spent Playing

Methods

In order to compare the amount of time spent playing video games during the exam week versus a normal week, let's analyze what we need to do. First, we should find the amount of time the students spent playing video games, which we can do by creating a summary of the 'time' column in our cleaned dataframe.

```
##   frequency   avg_time median_time count
## 1         1 4.62500000           2     8
## 2         2 2.61481481           2    27
## 3         3 0.06250000           0    16
## 4         4 0.04347826           0    23
```

In order to filter for the fact that the week was an exam week, let's make the assumption that many students that were studying for the exam were not playing games that week. In order to account for this, we filtered out students who played 0 hours of video games in the exam week. The reason we made this assumption is that it would mostly affect those that normally play daily or weekly, as compared to monthly or semesterly. If our assumption is that those who play monthly and semesterly would probably not even play, or play very little during an exam week, their distributions should remain relatively similar.

```
##   frequency avg_time median_time count
## 1         1 6.166667         3.0     6
## 2         2 3.069565         2.0    23
## 3         3 0.500000         0.5     2
## 4         4 1.000000         1.0     1
```

Now, let's use a boxplot to compare the two distributions, between students playing video games during exam week and those who didn't.

Frequency: 1=Daily, 2=Weekly, 3=Monthly, 4=Semesterly
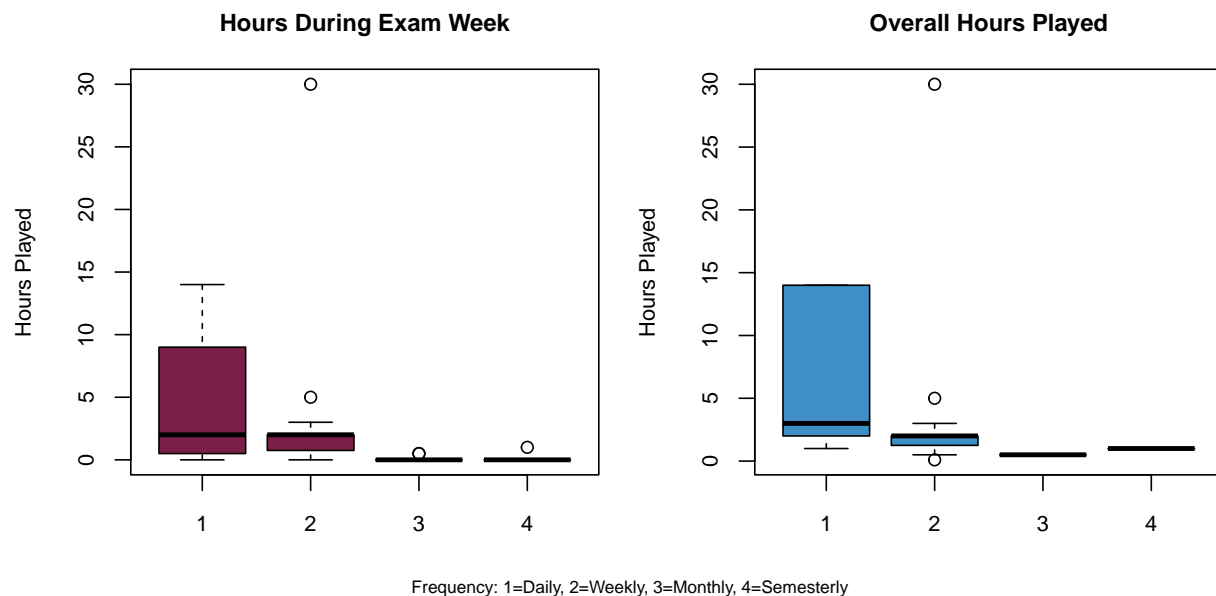
Analysis

Looking through these two boxplots and their data summaries, there is an obvious difference, especially for students with a daily or weekly frequency of play. For instance, our adjusted comparison shows that those who played ('freq' = 1) daily during an exam week typically had a lower amount of hours played, with a 1.5 hour difference between our exam week data and our filtered data. Similarly, our boxplot also shows a difference in the interquartile range between the two datasets for those who played daily, as during a normal week, students were more likely to play more and had a wider variety of playtimes, which can be inferred from the two box plots. A similar comparison can be made from students who played weekly ('freq' = 2). There's a ~0.4 hour increase in normal weeks versus the exam week. Another interesting observation is that

the adjusted data seems to have a smaller interquartile range compared to it's original exam week data, somewhat implying that those who play weekly usually play a more set amount of hours that usually doesn't diverge. Frequencies 3 and 4 don't have a large difference, just due to the nature of their descriptions, since most students in these categories would probably play very little to none already. If someone only plays games monthly or semesterly, then they are less likely to have major differences during an exam week versus a normal week.

Conclusions

Overall, the fact that there was an exam in the week prior to the survey tells us this:

- On average, students played less hours than they normally would on a week without an exam.
- Those who reported that they play daily or weekly typically have the largest difference between their average hours played during an exam versus a normal week, with around a ~1.5 hour difference for those who play daily and a ~0.45 hour difference with those who play weekly.
- Those who reported that they play monthly or semesterly have small or negligible differences between their average hours played due to the nature and scarcity of their data, along with the assumption that we made when filtering.

Question 3: Estimate Average Time Spent Playing

Methods

The point estimate for the average amount of time spent playing video games in the week prior to the survey is simply the sample mean. We will also calculate the standard error (SE) to represent the expected variability of our sample mean from the true population mean of time spent playing video games. This SE will also be used to construct our 95% confidence interval. The SE formula must use the **population correction factor** because we are sampling **without replacement** from a relatively small population of 314.

$$SE = \frac{\sigma}{\sqrt{n}} \times \sqrt{\frac{N-n}{N-1}}$$

[1] "point_estimate_average: 1.25977011494253"

[1] "Standard Error: 0.351851596413895"

Using the point estimate and standard error (in addition to an assumption of normality), we construct the 95% Confidence Interval.

[1] "lower_interval_estimate_average: 0.570153658068373 hours"

[1] "upper_interval_estimate_average: 1.94938657181668 hours"

To assess if our interval estimate is appropriate, we bootstrap 1000 sample means to estimate the sampling distribution. The mean and standard deviation of the bootstrap distribution are calculated.

[1] "Mean of bootstrap distribution: 1.27151034482759 hours"

[1] "Standard Deviation of bootstrap distribution: 0.416268625586636 hours"

Now, I construct a 95% bootstrap CI using the percentiles of my bootstrapped sample means and compare it to the earlier CI. If they are significantly different, the interval estimate may not be appropriate for this situation.

Bootstrap Sample Means with 95% CIs
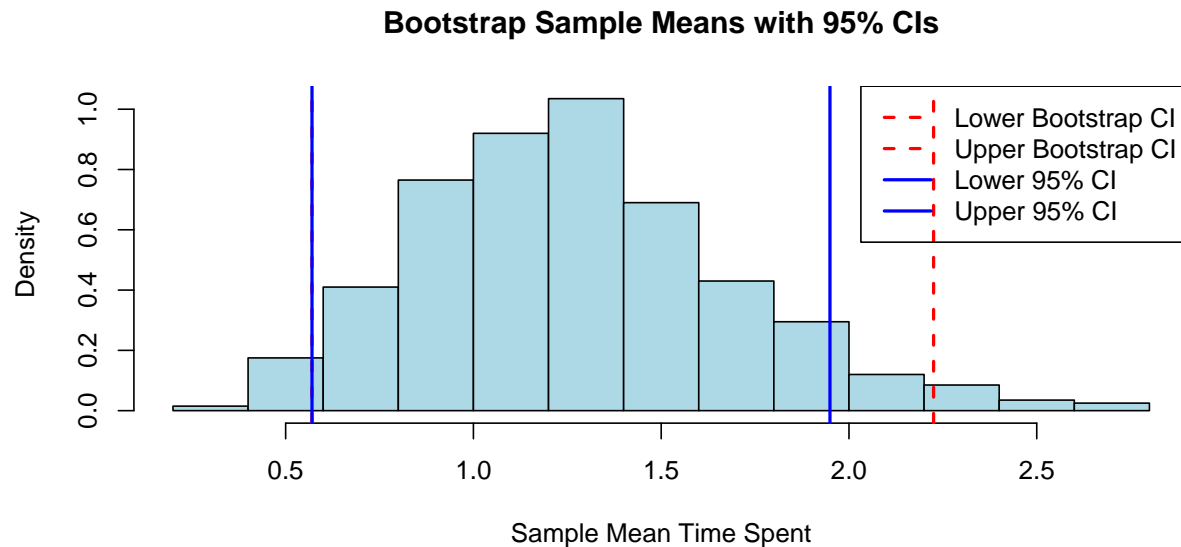
Analysis

Our 95% CI for the average time spent playing video games by UC Berkeley students was constructed under an assumption of approximate normality. It tells us that we are 95% confident UC Berkeley students play video games an average of 0.59 to 1.90 hours per week.

We bootstrapped 1000 sample means to estimate the sampling distribution. This bootstrap distribution had a mean of **1.23** and a standard deviation of **0.39**, while our original sample had a mean of **1.24** and a standard error of **0.33**. Additionally, the bootstrap histogram reveals a right skew. The difference between the sample's standard error and the bootstrap distribution's standard deviation in addition to the right-skewed shape suggest that the underlying population may not be perfectly normal. The graph also reveals a difference between the 95% CI and the 95% Bootstrap CI. The bootstrap CI extends further to the right than the initial 95% CI.

Conclusions

In conclusion, while our initial 95% confidence interval suggests that the true average time spent playing video games per week among UC Berkeley students is between 0.59 to 1.9 hours, there is evidence to suggest that this is not the most appropriate estimate to use. This is because the underlying sampling distribution of average play times might not be normal according to the shape of our bootstrap distribution. The variation in our bootstrap distribution is also noticeably greater than the standard error of our original sample, providing more evidence that the interval estimate is not appropriate.

Question 4: Important Reasons for Liking/Disliking Video Games

Methods

First, to get an idea of whether students generally like playing video games or not, we examine a bar plot of the "like" variable.

Counts of Responses to 'Like to Play' Video Games
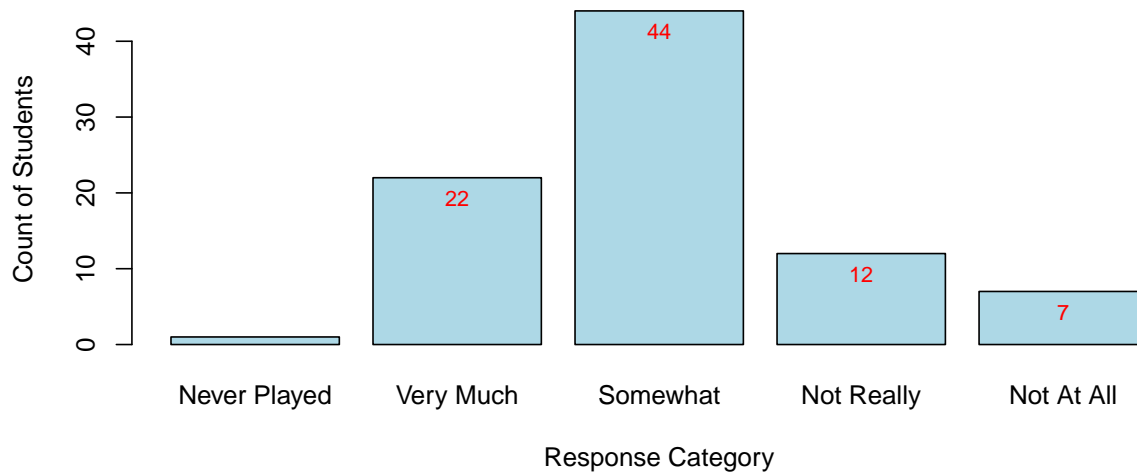
Additionally, we calculate the proportion of students surveyed that at least somewhat like playing video games, which turned out to be around 76.7%

Now, to find the most influential reasons for why students like or dislike playing video games, we use a bar plot to visualize the frequency of each reason.

Frequencies of Reasons for Liking/Disliking Video Games
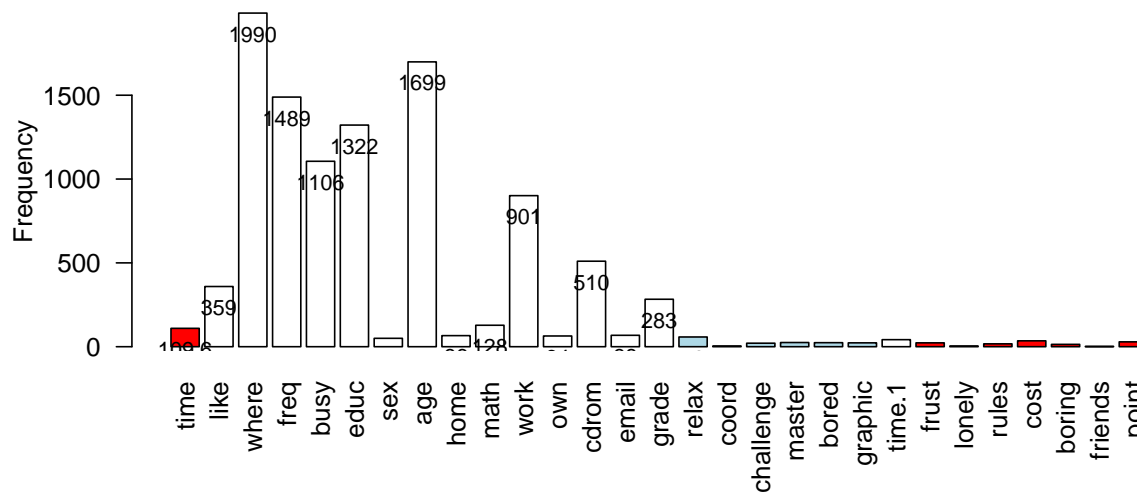
The most important reason for why students like video games is for relaxation. The most important reasons why students don't like video games are that they're time-consuming, costly, and pointless.

Analysis

Examining the initial bar plot of responses to the "Like to Play" question reveals that a majority of students surveyed at UC Berkeley like to play video games at least "somewhat". Only 1 student had never played video games and only 7 don't enjoy them at all. **76.7%** of students surveyed enjoyed playing video games at

least somewhat. The second bar plot displayed in this section shows the counts of reasons for liking/disliking video games as well as a color code representing whether the reason is positive or negative. The most common reason for liking video games is "Relaxation" with a response count of 58. The most common reason for disliking video games is "time consuming". The next most common reasons given are "Cost", "Pointless", and "Mastery".

Conclusions

In conclusion, the responses to the "Like to Play" question show that UC Berkeley students like to play video games in general, with **76.7%** of students responding with either "somewhat" or "very much". The most important factors for liking or disliking video games are "Relaxation", "Time Consuming", "Cost", and "Pointless".

Question 5: Compare Different Types of Students and Why They Play Games

Methods

To understand the differences between those who like to play video games and those who don't, we first collapse the range of responses to the last questions into 2 or 3 responses per question rather than 6 or 7. The responses to the question "Why do you play video games?" were placed into 2 categories: "Entertainment" and "Mental Engagement". The responses to the question "Why don't you like playing video games?" were placed into 3 categories: "Difficulty", "Low Priority", and "Isolation".

Next, we make cross-tabulations to show how various factors for liking or disliking video games change between different groups of people. The cross-tabulations graphs showing the most significant difference between groups are displayed below.
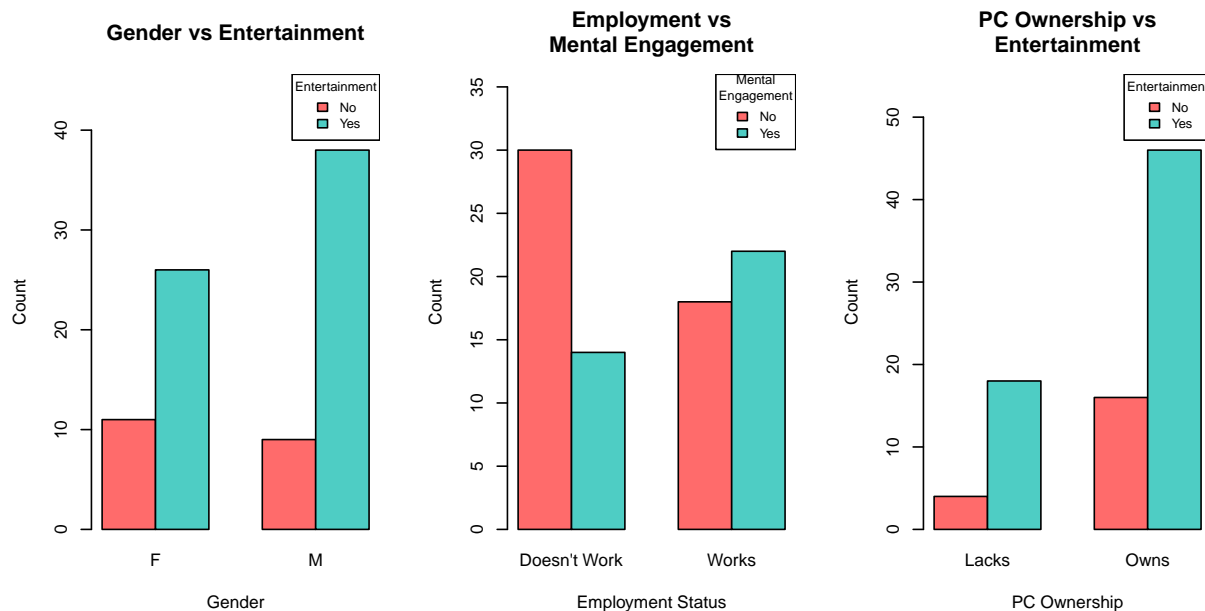
Analysis

Looking at our notable cross-tabulations reveals that, of those that responded to survey part 2, **70% of females** said they like video games for entertainment reasons in comparison to **80% of males**. This could suggest that male students are more likely to play video games for entertainment. Looking at the

second cross-tabulation reveals that **55% of students who work** for pay enjoy video games for mental engagement/stimulation, while only **32% of students who don't work** enjoy games for such reasons. This could suggest that students who work enjoy the mental engagement aspect of video games more. Finally, the third cross-tabulation reveals that **74% of students who own a PC** enjoy the entertainment aspect of video games while **82% of students who don't own a PC** enjoy the entertainment aspect. This insight may be surprising, but further analysis is needed to make a general statement about the population of college students. The rest of the cross-tabulations revealed much smaller differences between groups, suggesting that factors like gender, employment, and PC ownership don't strongly impact sentiments toward video games being difficult, isolating, and low priority.

Conclusion

In conclusion, the strongest contrast between groups of students detected in the survey was between students who work and those who don't. A significantly larger proportion of students who work enjoyed video games for mental engagement purposes, with **55%** for those who work and only **32%** for those who don't. The next strongest contrast is between males and females, with males being more likely to enjoy the entertainment of video games. Finally, students who own PC's in the survey were more likely to enjoy video games for entertainment than those who don't.

Advanced Analysis

Our advanced analysis focuses on the intersectionality between people that like video games and which types of video games they prefer. By doing this, we can see the most popular types of video games that these students prefer, along with the average amount of time they play video games a week, potentially showing us which games people spent the most time on.

Methods

In order to do this, we filtered the data. It's important to note that we ignored people who skipped this question, because those aren't the people we want to analyze. We went with the assumption that those who put all 0s in each of the columns that corresponded to the game type, because presumably if someone didn't like games or haven't played it, they wouldn't know the genres and skip.

```
## [1] 75 34
```

From this, we can see that we have 75 respondents that still enjoy games, which we can further filter out. WE can do this by capturing the most popular video game genres from this data, while also comparing the amount of time people spent playing those games.
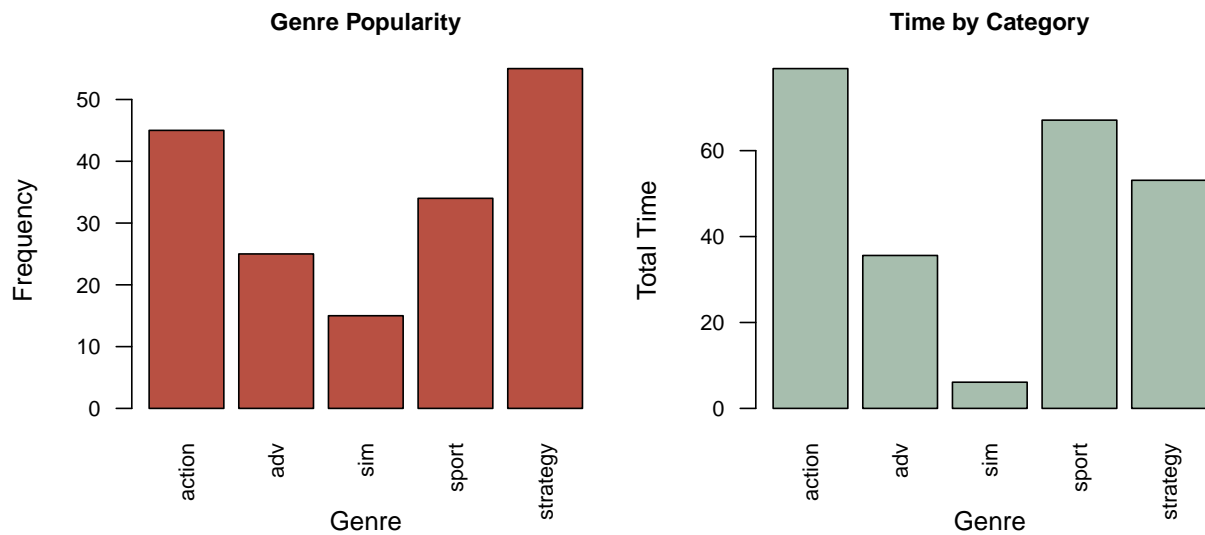
Analysis

Our two graphs, especially when put next to each other, show us another interesting detail. Although strategy was the most popular game genre chosen by people, with the action and sport genres falling slightly behind, when we look through the total time spent playing each game, action was the highest, with sport and then strategy falling behind. This means that there's a decent chance that the people who play strategy games don't typically play the game for a longer time, while those who played action and sports games play for longer durations. On another note, sim was the least popular genre of game in the survey, and that pattern still shows in the total time spent. People who mentioned they like simulation games spent very little time actually playing them, leading us to believe that those who play these simulation games either don't play them very often or play them for short periods of time compared to their peers who played action or sport games.

Conclusions

Our conclusions for this advanced analysis show us that although a video game genre like 'strategy' could perhaps be the most accessible, it's not necessarily the most played time-wise. While 'action' and 'sport' genres were not as popular as 'strategy' games in terms of total picks, the students who played these games far exceeded the play-time of students who played 'strategy' games.

Conclusions and Discussion

Summary of Findings

Overall, we found a lot of interesting things about our data. First of all, one major thing we noticed is that a significant portion of students, even during an exam week, played video games. Our own assumptions also led us to believe that students would probably be more likely to play more video games when there was not an exam week as well, suggesting video games may be a way to diffuse stress or relax. We also found interesting similarities in the average amount of time students spent playing video games, typically over an 1 hour per student per week. We also found that most students liked playing video games, specifically

within the action, sports, and strategy genres, though action and sports were played the most (in terms of time). Additionally, we also found that students typically played video games for relaxation purposes, while those who didn't cited complaints that video games were too time consuming, costly, or pointless. We also analyzed reasons as to why students don't play video games or what factors may differentiate students. For example, those who didn't work typically had didn't cite mental engagement while playing video games as a reason why they played video games, while those who work did. Another factor was gender, as males typically played games more for entertainment than females. These shows signs that people typically play games for different reasons, and there are a lot of intrinsic and confounding factors that one has to take account when doing such an analysis.

Discussion

Our EDA on this data was important in the way that it related the enjoyment and usage of video games from college students, along with the reasons why people either enjoyed/disliked them. This is relavent towards the industry as a whole, because many game companies want to learn what draws people to games, as it's a big selling point. However, our study is nowhere near conclusive. We only had 89 responses out of around 300 people, and this is an extremely small population compared to something a company would want, meaning that our data is probably not a good simple random sample that could represent a whole population. However, this EDA helped analyze potential reasons and put a small correlation as to why general people might play video games.