Identifying the Potential Correlation between Computer Configuration and Mental Discomfort Using Logistic Regression

Junyuan Quan (Derek)
Research Advisor: Dr. Matthew Lavin
April 22nd, 2023

Abstract

This research investigates the correlation between computer configurations that are recommended by video games and mental discomfort that players experience when playing video games. The goal is to determine whether video games with greater requirements are more likely to cause mental discomfort or not. To accomplish this, the text data from player reviews on Metacritic and the recommended system requirement of video games from Steam were used. An in-depth analysis of logistic regression was employed to estimate the probability of feeling mental discomfort with different combinations of computer configurations. The analysis revealed a positive correlation between the amount of required computer storage and mental discomfort, meaning that games with higher storage requirements are more likely to cause people to feel uncomfortable. On the other hand, a negative correlation was found between the release year of games and mental discomfort, indicating that newer games are less likely to cause mental discomfort. With additional data, it will be possible to identify specific combinations of computer configurations that do not cause mental discomfort for players. The findings of this research could provide recommendations for players about the guidelines for selecting enjoyable games and for game developers on how to optimize game configurations to reduce the risk of causing mental discomfort for players, while still maintaining visual appeal and popularity.

1 – Introduction

1.1 – Background

The video game industry is one of the fastest-growing industries. The genres and platforms of video games have evolved dramatically during the past decades. This evolution has allowed the public to have access to video games in various ways and in different environments. It has become one of the most popular ways to relax after intensive work or study. With the progress of technology, game manufacturers have greatly developed image quality – the pixel games (2D) of the very first generation have progressed to high-definition (HD) games (3D) and virtual reality (VR). The rapid evolution of image quality attracted an increasing number of publics to play. However, many individuals have reported visual discomforts like dizziness and headache after long periods of time playing video games (Zhou, Jun, et al., 2019). Some of these symptoms may have further consequences such as poor sleep quality (Akowuah, Prince Kwaku, et al., 2021). In response, some adults may recognize the impact of video games on their mental health and adjust their gameplay accordingly. At the same time, many others may not have the same self-discipline and continue harming their mental health. Therefore, having a greater sense of which games are more likely to cause mental discomfort may help to minimize the harm to the eyes and mental health from video games.

The main focus of this research is to investigate the underlying correlation between computer configurations like operating system, processor, graphics card, memory, and storage and players' mental feelings like dizziness, headache, tiredness, and eyesore, which should be concerned by both players and game developers since players may maintain a safe health status when playing enjoyable video games while game developers may improve or adjust the default game resolution to minimize the harm to players' mental health, accumulate popularity, and

make broader profits later on. The analysis of this research is done based on the data from players' comments from Metacritic, an online source that compiles reviews of entertainment content, and recommended system requirements of video games from Steam, a platform that can be used to browse PC games. This research conducted a thorough data analysis using logistic regression to calculate the probability of feeling mental discomfort across different combinations of computer configurations. The hypothesis is that games that require better computer configurations, including better processor, graphics card, etc. are more likely to cause mental discomfort among players since better computer configuration typically means higher resolution and frame rates, and higher resolution and frame rates typically means faster motions and changes on screens. This does not imply that games with fewer requirements are more enjoyable because these games may not be challenging and playable but are tedious, so there is no sufficient tradeoff for players to give up games with better requirements. Another hypothesis is that games that are released in recent years are more likely to cause mental discomfort symptoms because technologies of making video games develop as time goes, and better technology always motivate game developers to create video games that are more realistic and magnificent. The hypotheses might not be true if insufficient data is used in analysis and inappropriate benchmarks are used to quantify the system requirements. The models are compared in the later process through accuracy and recall per class to estimate the performance of the models. The model with the best performance is chosen to interpret the results and make the final recommendations.

The next subsection displayed a literature review that identifies the existing research in the area and the motivations of this study. The differences between computer configurations, the effect of visual display terminals like computers on players' visual status, and the potential approaches to reduce mental discomfort that were discovered by previous researchers were

reviewed to refine the goal of this research. The last subsection mentioned the ethical considerations that this research may encounter. In section 2 the data collection process as well as the approaches adopted to clean the data are described, including the transformation of numeric data from text. The methods including the models used are also elaborated in this section to provide the basis of knowledge of data analysis in this research. In section 3 the descriptive results including visualizations and modeling output are displayed. In section 4 the interpretations of the results and their implications are elaborated. And lastly, section 5 discussed the limitations of this study and some interesting works that have not been done in this study but may be developed in the future. Since mental discomfort relies on visual display from computer screens, the results and decisions from this research could be regarded as recommendations for players of choosing more enjoyable video games and for game developers to adjust computer requirements to secure video game player health and well-being.

1.2 – Literature Review

1.2.1 The differences among various video game settings

Video games can often be played on different platforms with various display capabilities, like a personal computer (PC), Xbox, Switch, PlayStation, and VR. This diversity of platforms results in different frame rates and resolutions. Previous studies have investigated the effects of frame rate and resolution on prevalent games. Claypool, M and Claypool, K (2009) have found that frame rates have a much greater impact on user performance than resolution, where the frame rate is typically the frequency at which consecutive images are captured or displayed (Brunner, 2022) and resolution indicates the number of pixels that are displayed per inch for an image (Awati, 2022). In particular, player performance shows a significant drop in performance

below 15 frames per second within a modest increase from 15 to 30 frame rates per second. Further, player performance is not affected by resolutions from 800x600, a low-end computer setting, up to 1280x1024, a higher-end setting. These results might be valid and reliable back in the day since 1280x1024 might have been the best resolution at that time, however, with the rapid growth of technology, frame rates, and resolution have developed greatly in the past decade. The situation is different for the latest hardware and platform, 10 may be the lowest frame rate and 1280x1024 may be the highest resolution in the past, but the frame rates can be as low as 1 and resolution can be up to 3840x2160 (4K) today. Besides player performance, Claypool, M and Claypool, K (2009) also mentioned their analysis of player playability ratings versus frame rate and resolution, and it shows that resolution was a bigger factor for playability than the frame rate. With both of their findings, it can be concluded that frame rate, resolution, and other computer or game settings indeed have effects on players' feelings. The later part of this section will discuss the different effects that video games and computer screens have on players' mental health.

1.2.2 The effect of visual display terminals on players' mental health

Visual fatigue and visual discomfort are in fact two distinct terms. They are both diagnostic terms for asthenopia. Asthenopia may be concentrated around the eyes or may further cause general headaches or occur in the neck and shoulders. The specific distinction was made by Lambooij, Marc, et al. (2009) that visual fatigue refers to a decrease in the performance of the human vision system, which can be objectively measured, while visual discomfort is its subjective counterpart. According to Zhou, Jun, et al. (2019), the visual brain fuses the left and right images projected onto the two eyes from a stereoscopic 3D display, perceives parallax, and

rebuilds a sense of depth. In this process, the eye adjusts vergence and accommodation to adapt to the depths and parallax of the points they gazed at, which causes visual discomfort with the conflict between accommodation and vergence when viewing 3D content. Similarly, it is also stated by Lambooij, Marc, et al. (2009) that temporarily changing the demand of accommodation-vergence linkage, for example, by fast motion in depth can cause visual discomfort. Further, three-dimensional artifacts resulting from insufficient depth information in the incoming data signal yielding spatial and temporal inconsistencies, as well as unnatural blur are also underlying factors that cause visual discomfort. Visual fatigue is generally measured with optometric devices. However, those measurements on the visual system are normally costly, time-consuming, and only able to be conducted on a small number of participants, which makes the results not reliable (Lambooij, Marc, et al. 2009). With this limitation, it seems hard to monitor potential damage to the visual system as a result of prolonged viewing. On the other hand, clinical measurement methods seem more appropriate in this situation for their feature of relatively cheap, concise, noninterventional, quantitative with high sensitivity and specificity, and applicable to a large group of participants (Scheiman and Wick, 2008). In fact, the factors that cause visual discomfort are essential characteristics of video games. Most of the latest video games are known as lifelike and immersive. For example, the latest first-person shooter games make players feel like holding real guns and fighting in the field, whereas survival role-playing games make players feel like they are living in prehistory and striving for survival. Besides these games on personal computers, such types of games in VR are far more realistic. However, the factors mentioned above are common in today's video games, which makes them undoubtedly harmful to players' visions.

With the recognition of the harmful characteristics of visual display terminals, recent studies have investigated the reasons and effects of playing video games among adolescents. Video games according to Eskasasnanda (2017), are popular among students because adolescents seek self-identity and independence, and video games are the platform to achieve that goal. On the other hand, video games are sophisticatedly designed to amaze and evoke the players' curiosity with their attractive game plays, realistic images, and sounds. Eskasasnanda's study (2017) was conducted by interviewing high school students in Malang City, East Java Province. In his interviews, some students responded that playing video games too much sometimes makes them hard to stop since the realistic image and sounds are designed to challenge players. Furthermore, playing too many games makes students hallucinate. Some students also reported that video games make them physically exhausted, and they experience dizziness and back pain because of too much sitting and watching computer screens. All these negative impacts on students can be concluded that they are caused by long time sitting and long-time watching screens. However, with the fact that long-time reading instead of watching screens won't cause hallucinations and addictions, computer screens seem to be the primary cause of these harms. Therefore, understanding computer configuration and resolution is important to ascertain the source of visual discomfort and mental issues like hallucination and addiction.

1.2.3 Potential approaches to reduce mental discomfort.

Previous paragraphs have discussed the harmful effects of computer screens on people's mental health. However, since the progress of technology continues to flow forward, computer screens will still be prevalent as the main medium of video games in the foreseeable future. Therefore, finding a way to reduce the harmful effects will be a potential focus for game

developers to improve games and achieve greater reputations since games will become more enjoyable. Hoffman, David m., et al. (2008) mentioned that they have developed a novel 3D display that presents focus cues that are correct or nearly correct for the depicted scene, which causes the time required to identify a stereoscopic stimulus to be reduced, stereoacuity in a time-limited task to be increased, distortions in perceived depth to be reduced, and viewer fatigue and discomfort to be reduced. These benefits correspond to the factors that cause visual discomfort in the previous section, and it proves that it is possible to avoid or reduce negative effects that are provided by high game resolutions from computer screens.

Besides well-known harmful effects brought by computer screens and video games, recent studies and the latest technology have developed some video games to help adults with dizziness to recover. It is found by Viirre and Sitarz (2009) that immersive computer environments can improve vestibulo-ocular reflex (VOR) function and reduce vertigo. Phillips, et al. (2018) have conducted a single-site prospective clinical trial on patients with dizziness. They found that video gaming using the Wii Fit balance platform can be used for balance rehabilitation and it is a cost-effective way of exercising. Furthermore, the application of vestibular rehabilitation in a virtual reality (VR) environment can lead to additional improvements especially in dizziness symptoms, disability, balance, and mobility in the elderly with chronic dizziness (Kanyilmaz, et al., 2022). Based on what we have known before and what are showed in this paragraph, it appears that computer screens and video games can either be harmful to people's vision and mental health or be beneficial to people that already have mental illnesses. This is contradictory but at the same time makes sense because video games are designed for different purposes. But it motivates curiosity about the boundaries between these two types of games.

This section evaluated the possible correlation between video games and players' mental health. The first part pointed out that various settings of video games and computers can have different effects on players' feelings. The second part of this section explained the cause and effect of visual fatigue and visual discomfort, as well as the reasons that video games are popular among adolescents and their influences. The third part of this section presented previous studies that explained possible ways to reduce visual fatigue and visual discomfort and some of the applications with recent technology. However, since computer screens can be both harmful and beneficial with different applications, it may be essential to ask about the boundaries between the two different types. From my personal experience and responses from players after long gameplay time, it is common that people become tired as more time is devoted to playing video games, which makes it true that those video games that are designed to help patients recover are essentially different from the commercial games. Therefore, the goal of this research is to investigate the underlying differences between these two different types of video games, in other words, the different impacts on mental health among different computer configurations like operation systems, processors, memory, graphics card, and storage requirement. If there exist some configurations that make players feel more comfortable, it will improve players' experiences to be more enjoyable. At the same time, if game developers could improve games in the same way, they can win greater reputations and secure players' mental health and well-being.

1.3 – Ethical Considerations

In this research, players' comments are used as text data and were transformed into numeric values. The participants are the individuals who have left comments on Metacritic and may have shared their personal information or opinions about their mental health. Therefore, one

possible ethical consideration would be to ensure that the data displayed on Metacritic is de-identified and that any identity information is removed or hidden. Fortunately, all users on Metacritic use their usernames, which are not necessarily to be their real names except some users may use their real names. In this case, it becomes impossible to recognize the true identity of all the users. Furthermore, the data obtained from Metacritic does not contain usernames, so this ethical consideration is not a concern.

Another ethical consideration is the potential bias in the logistic regression models in this study. In the data cleaning process, all the system requirements data were converted into numeric values, some of them were nominal numbers from original strings, and some of them were mapped with benchmark ratings from online sources. Possible bias may include that the performance or effect of 2GB RAM may not be twice of 1GB RAM, but the data been used is straightly 1 and 2 for each game. This may be addressed with additional reliable sources. Furthermore, since some of the games either did not have complete requirements for computers (only one or two requirements) or their hardware was too outdated and was in an unrecognizable format, these missing values were filled in with either "1" or the average values of low-end hardware's benchmark, which could raise the issue of bias because they may conflict with existing values (high requirement of graphics card but processor value is filled in with a small value). Besides, keywords were used to filter comments that express players' mental discomfort. However, it might be the case that players refer to the word for the opposite meaning, for example, the complete comments may have the phrase "not tired", but my code regarded this kind of comment as expressing mental discomfort, which could also be a possible source of bias.

Finally, the potential implications of these research findings for players and the broader gaming community should be considered. If the results suggest that certain computer

configurations are associated with negative mental outcomes, it is important to communicate these findings in a sensitive and responsible manner, and it is important to provide recommendations for players and game developers to mitigate the negative effects on mental health.

2 – Data & Methods

2.1 – Data Acquisition

2.1.1 Metacritic comments

The first set of data comes from players' comments for video games on PC, and the comments are from Metacritic, one of the largest websites that compiles reviews of entertainment contents. The focus of this study is only video games on PC because game requirements from other platforms are harder to obtain and video games on other platforms require extra devices. The total number of games that are rated, meaning that they have at least some numbers of comments and reputation, on this page is 5,321, and all the comments for the games are included in this part of data.

The data is collected through the following steps: 1) iterate through all the 54 pages in this website and obtain the name of games on each page, 2) convert the names into URLs for each game by simply plug in the names into the original website and end with "user-reviews", 3) obtain the number of comment pages for each game since different games have different number of comments, and put the game name, corresponding URL, and total number of comment pages altogether. 4) Scrape comments by iterating through each game and iterating through all the comment pages each game has. The Python package "BeautifulSoup" is mainly used for scraping

this set of data. This package traverses the HTML tree structure of web pages and allows access to specific HTML tags, attributes, or text content to extract the relevant data.

2.1.2 Recommended system requirement from Steam

The second set of data comes from Steam, the largest platform for PC games. The system requirement for each game is obtained, which typically include OS (operating system), Processor, Memory, Graphics, DirectX, and Storage. Most games have both minimum system requirements and recommended system requirements, and the recommended ones are considered in this study because it is assumed that most players have even better computer configurations than the recommended requirements. The release date for each game was also collected from Steam, as it may be helpful for future analysis. The original attempt was to obtain data for games that are included in the first set of data, but since Steam requires the App ID of each game for access, the APP ID was collected from a third-party database that provides information on Steam games (SteamDB). The games that are available in this database are scraped, and the total number of games in this database is 6,081.

This set of data started with the collection of APP IDs. Since SteamDB is poorly scrapable, the Python package “Selenium” was used instead of “BeautifulSoup” in scraping the name of games and their corresponding APP ID, where “Selenium” simulates user interactions with web pages. After obtaining the game names and corresponding APP IDs in lists, the data was converted into URLs to prepare for scraping information from Steam, following the same process as the first set of data. While some of the URLs may not work for their problematic format or returning to the home page of Steam, about 5,257 games were successfully scraped in the end.

2.2 – Data Cleaning

2.2.1 Metacritic comments

The raw data of the comments from Metacritic cannot be used for analysis directly due to their unstructured format and the presence of unreadable symbols. To quantify the comment texts, a list of keywords, including dizzy, dizziness, tired, tiredness, eyesore, headache, and fatigue (upper case formats are considered) is used. Comments containing any of these keywords are considered to express mental discomfort. After aggregating comments for each game, the number of comments expressing mental discomfort is calculated, resulting in either zero or an integer value. Additionally, the percentage of comments expressing mental discomfort is computed by dividing the number of relevant comments by the total number of comments for each game. Since this study aims to identify the relationship between computer configuration and mental discomfort, specifically which types of games are more likely to cause such discomfort, the percentage of comments expressing mental discomfort is converted to a binary variable, where “0” indicates the game does not cause mental discomfort, and “1” indicates it does. If the percentage is greater than 0, it is converted to “1”, which applies to less than half of the data.

2.2.2 Recommended system requirements

The raw data of recommended system requirements also cannot be directly used for analysis since they are also text data and many of them have symbols that need to be decoded. To prepare the data, the raw strings are first separated into multiple parts: OS, Processor, Memory, Graphics, DirectX, and Storage, and each part is assigned to be a new column. While most games have proper format and are separated successfully for later use, some games have

different format, and the requirements for these games are either considered missing or filled in manually.

Processor and graphics requirements typically specify specific brands, while operating system, storage, and memory are loose requirements, such as “1GB RAM”, “Windows XP/7/8”, and “200MB available space”. To quantify the brands required by games for processors and graphics, benchmark ratings from UserBenchmark are used, where the benchmarks are effective speed that calculated based on gaming performances for the top 12 popular games. The website rates 1,405 CPUs (processor) and 698 GPUs (graphics card) and covers most brands and models in the data. By mapping the models with the average benchmarks provided on this website, a benchmark score is assigned. A score of “2” is assigned for games with missing graphics requirements and “27.7” for games with missing processor requirements (estimated average benchmark for low-end CPUs and GPUs).

For the other three requirements, since they are not specified as a particular brand or model, it is unlikely to find any benchmark to quantify them. For memory and storage, these requirements are converted into numeric values straightly from themselves, though it should be noted that 2GB memory is not necessarily twice as good as 1GB memory. For games that have missing storage and memory requirements, a value of “1” is assigned to each. For operating system, since there are only a few categories, including Windows XP/7/8/10/11 and 32/64 bits, a value of “1” is assigned for the basic Windows XP, with an additional unit added for each subsequent OS requirement level. For example, “2” is assigned for Windows 7, and “3” for Windows 8. If the requirement specifies at least 64-bit system, an additional unit is added to the value. Games with missing operating system requirements are also assigned a value of “1”.

The two sets of data are finally joined together after the data-cleaning process on each. Since the name of games from the two datasets may not be in the same format, all the symbols and spaces are removed between each character but there are only 1,439 games are successfully matched. After clearing out the games that have missing values for all the five requirements, the final data have 1,349 rows in total and the sample data looks like below:

Game	Year	Discomfort	OS	CPU	GPU	RAM	HD
Assassin's Creed Valhalla	2022	1	6	71.4	32.9	8	160
Forza Horizon 5	2021	1	5	81.8	79.7	46	110
Destiny 2	2019	1	6	60.6	49.1	8	105

Table 1: Sample data of the final dataset

The variable "Year" is the release year of the game, "Discomfort" is the binary variable that indicates whether the game cause mental discomfort or not, "OS" is the value for operating systems, "CPU" is the benchmark value for processors, "GPU" is the benchmark value for graphics cards, and "RAM" and "HD" are the nominal number for memory and storage (hard drive space).

2.3 – Methods

Logistic regression was the primary statistical classification method used in this study to analyze the relationship between a binary outcome variable and predictor variables. As mentioned in the previous part, a new column of data was created that converted the percentage of comments expressing mental discomfort into a binary variable, indicating whether a game is likely to cause mental discomfort or not based on whether it had comments mentioning certain

keywords. The goal of the model is to calculate the probability of the outcome variable taking a particular value as a function of the predictor variables. In binary classification methods, evaluation metrics such as a confusion matrix, accuracy, and recall per class are typically used. A confusion matrix summarizes the performance of the classification algorithms, accuracy measures how often the algorithm makes correct predictions, and recall per class measures the proportion of actual positive instances that are correctly identified by the algorithm for each class, in this case, whether a game causes mental discomfort or not.

To use logistic regression effectively and have trustworthy outcomes, some parametric assumptions need to be met: 1) linearity, 2) no outliers, 3) randomness, 4) independence, and 5) no multicollinearity. Linearity ensures that the relationship between the explanatory variables and the log odds of the response variable is linear. Normality refers that the residuals of the logistic regression model should be normally distributed. Randomness means the data is collected randomly. Independence indicates that the game system requirements and comments collected are independent of each other, that the values of one game do not affect others. And multicollinearity refers that the explanatory variables having no linear relationships. However, logistic regression may still be useful and effective even if not all these assumptions are met, as there may be other factors to consider depending on the research and data being analyzed. Additionally, since the goal of this study is to find which or which set of system requirements are more likely to cause mental discomfort, it is important to recognize what variables are significant and contribute the most to the model, which requires a rigorous interpretation of the model summary and evaluation metrics.

3 – Results

Based on the standard of creating the binary variable mentioned earlier, 785 games do not cause any mental discomfort and 564 games are likely to cause discomfort symptoms. First, the linearity assumption of the explanatory variables was checked using empirical logit plots to see if the variables need transformations to perform better in models.

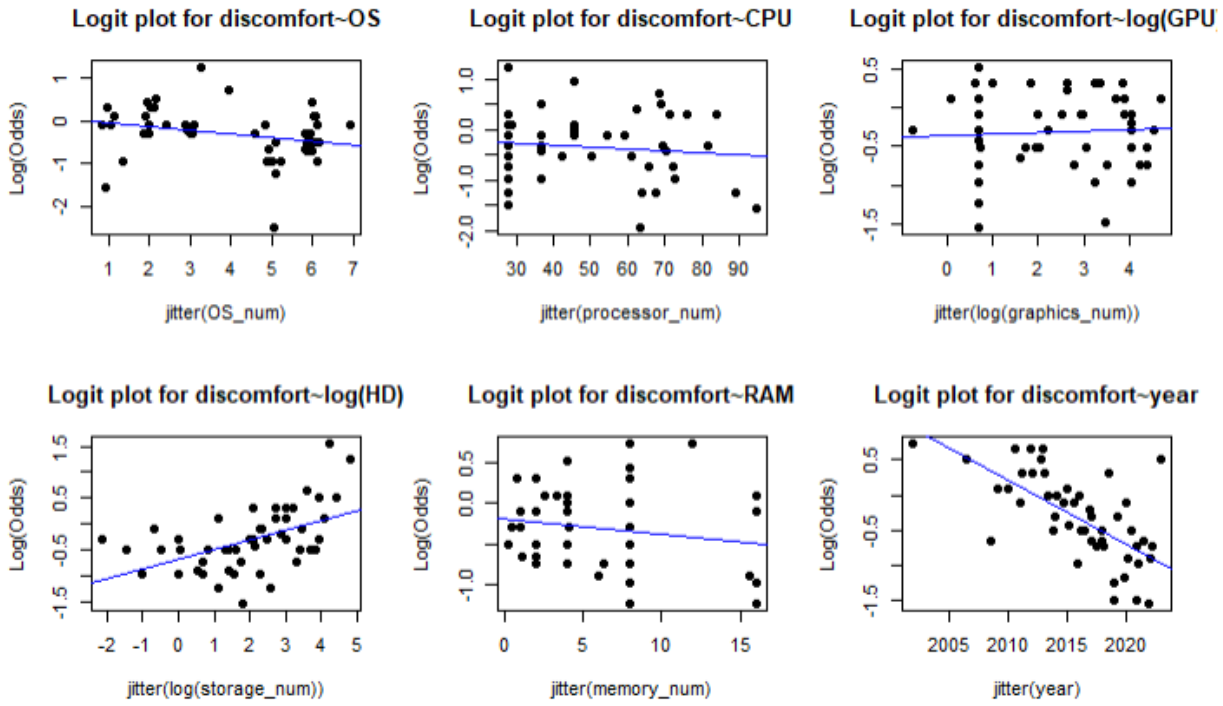


Figure 1: Empirical logit plot for OS, processor, graphics, storage, memory, and release year

The above figure displays the empirical logit plot that shows the linear relationship between the natural logarithm of odds and independent variables. Among the variables, storage and graphics are log-transformed because groups of data are skewed on the original plot (See Fig. 2 in Appendix), and the transformed data may be more significant when fitted in the model. Based on the figure above, it appears that all the variables satisfy the linearity assumption, and it can be assumed that OS, year, and log-transformed storage are important in the model while the

remaining variables are less significant because their slopes on the plot are close to 0. By fitting logistic models with different selections of variables (with choices of including year variable or not, including log transformation or not, and including interaction terms or not), the following table displays the accuracy and recall per class of each model:

	sys_req_mod	year_mod	trans_mod	interact_mod	both_mod
Accuracy	0.616	0.630	0.660	0.665	0.668
Recall for discomfort	0.207	0.386	0.452	0.488	0.506
Recall for not discomfort	0.914	0.805	0.810	0.792	0.784

Table 2: Summary statistics of fitted logistic models

The latter models build on the former models. The sys_req_mod is the model with only the five system requirements variables involved without any transformation on variables, the year_mod adds the release year of games to the sys_req_mod, the trans_mod transformed the storage and graphics values, the interact_mod does not transform the two variables but adds the interaction terms between each variable, and the both_mod is the model with all the variables, transformed variables, and interaction terms. A higher accuracy means more correct predictions are made on the test dataset, a higher recall for discomfort means more games that cause mental discomfort are correctly identified, and a higher recall for not discomfort means more games that do not cause mental discomfort are correctly identified. More confusion matrices are displayed in Appendix.

To evaluate what variables are significant in the model, the variable importance plot and the table of coefficient scores and p-values of the top 6 most important variables are shown below:

	trans_mod			both_mod	
Variable	Coefficient	p-value	Variable	Coefficient	p-value
year***	-0.143	2.64e-10	storage_memory	1.82e-3	0.2106
storage***	0.321	1..71e-7	year	-1.32e-1	0.0778
graphics	0.117	0.178	graphics_OS	5.73e-3	0.1671
memory	-0.015	0.551	storage_graphics	8.28e-5	0.7537
OS	-0.028	0.607	OS_memory	-3.08e+1	0.2909
processor	0.002	0.713	storage_processor	-4.22e-4	0.1972

Table 3: Coefficients and p-values for the top 6 important variables in the two models

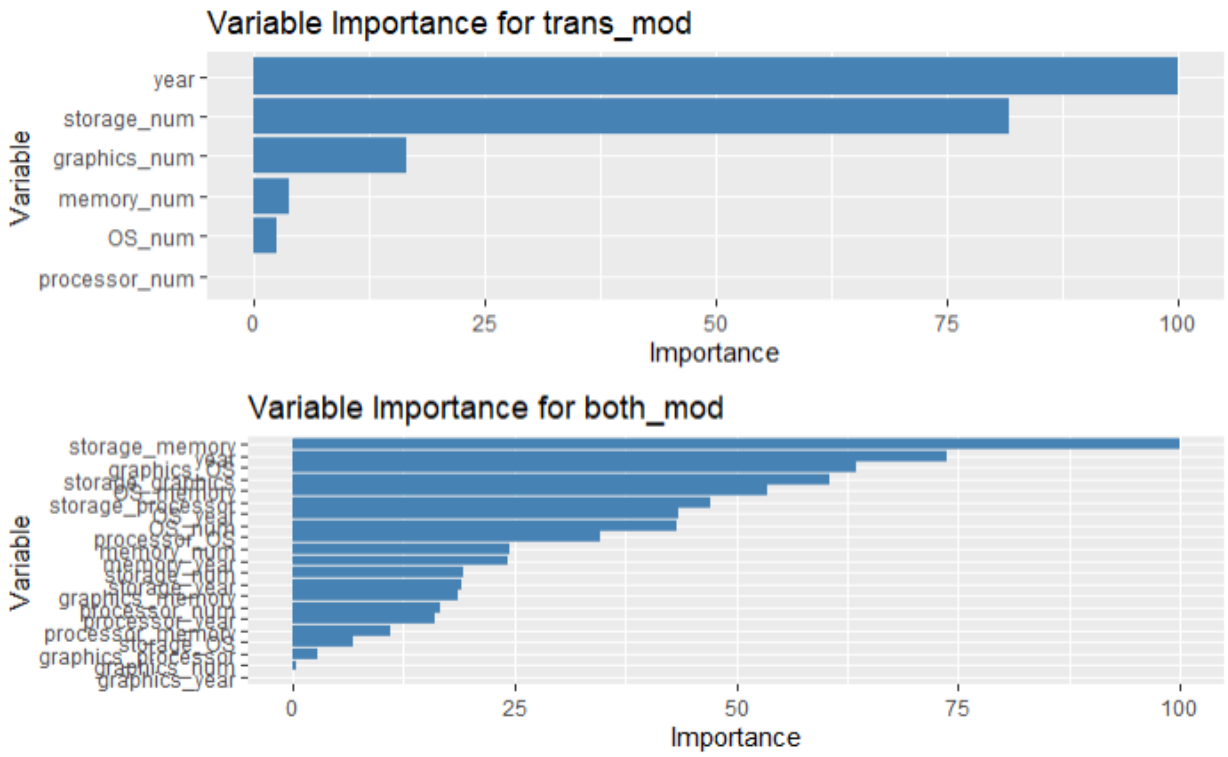


Figure 2: Variable importance of the trans_mod and both_mod

The trans_mod and both_mod are selected for variable evaluation because trans_mod is less complicated with only the variables from the data while both_mod may present more details and

stories relative to mental discomfort. The larger bar in the variable importance plot means the variable is more important, and the stars "*" in the table represent that the variable is statistically significant in the model and more stars mean the variable is more significant. Table 3 shows that the year and storage are significant in the model with transformation (trans_mod) while none of the important variables in the full model (both_mod) are significant.

The model with transformation terms is further examined because it has relatively decent accuracy and recalls per class and at least some of the variables are significant. The relationship between each explanatory variable and mental discomfort is further plotted with overlaid logistic regression line in the Appendix. The fitted logistic regression line for processor, graphics, memory, and OS are nearly horizontal lines centered around 0.5, which further verified that they are not likely to be significant factors of mental discomfort.

The other assumptions of the model with transformation are also checked. For multicollinearity, none of the VIF values are higher than 4, indicating that none of the explanatory variables are highly correlated with the others. The outliers are checked through the Cook's distance plot and plot of residuals vs. leverage values:

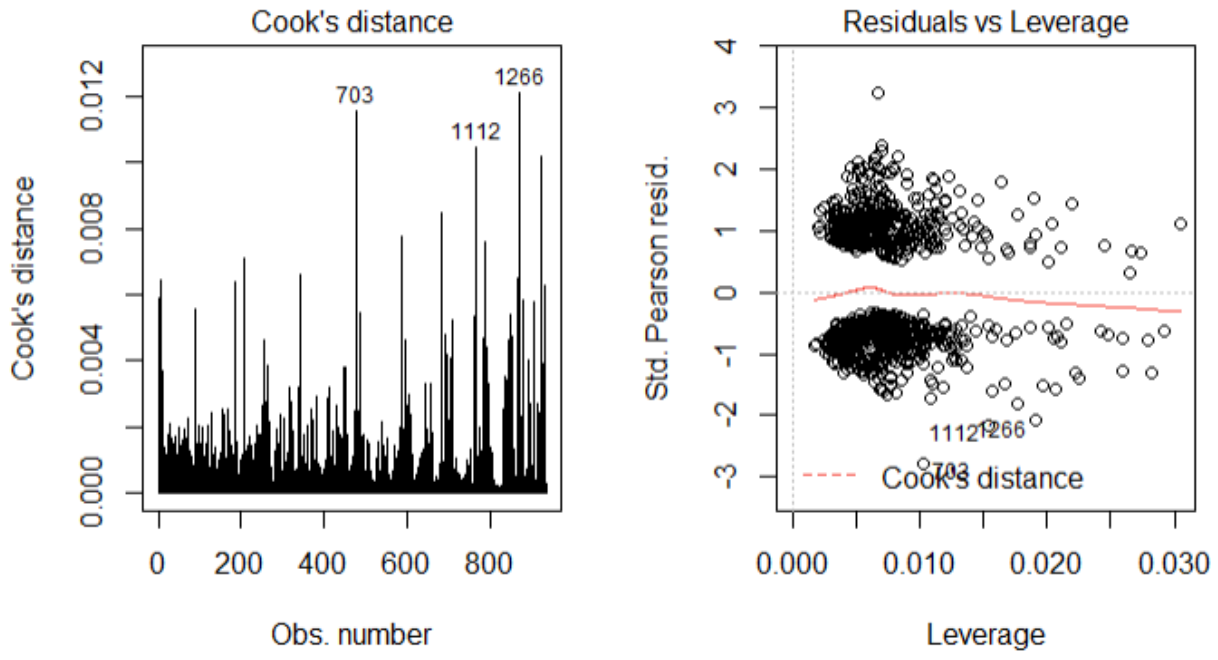


Figure 3: Plots to check for potential outliers.

It appears that there are three outliers in the data with outstanding Cook's distance displayed on the Cook's distance plot. However, the residuals vs. leverage plot suggests that the three outliers are not significant (within the Cook's distance), that they do not have high leverage so they do not have a large influence on the model fit. Therefore, no outliers need to be concerned. For randomness, since all games are collected from two websites, the entire data can be considered as the whole population. However, since only a fraction of the data is combined because of some formatting errors, the final data can be thought to be random. Furthermore, the training and testing sets are separated randomly from the final data, so randomness should not be a concern. For independence, on the one hand, due to the social nature of commenting and the potential for players to be influenced by others, there may be some concerns about the independence of the comment data. On the other hand, the system requirements of games are independent because they only depend on hardware itself.

4 – Discussion

The major objective of this research is to find the relationship between mental discomfort and computer configuration, in particular, which requirement of games is more likely to cause mental discomfort. The hypotheses are that 1) games with higher system requirements are more likely to cause mental discomfort and 2) newer games are less likely to cause mental discomfort. These hypotheses might not be true based on some hidden facts about video games. On the one hand, mental discomforts like dizziness and vertigo are not linked to particular games but to how players behave, such as the length of gaming sessions, lack of water, posture, etc. On the other hand, there are many games in the data that may cause mental discomfort, but the players do not report it. Additionally, the "field of view" in games may also affect players' feelings about games but is not included in the data. The field of view refers to the extent of the observable environment that a player can see on the screen. A larger field of view means that the player can see more of the game world around them, while a smaller field of view provides a more focused view. The field of view is determined by the game itself, and it can have impacts on player immersion, perception, and gameplay experience. Previous studies have shown that it is difficult to quantify people's feelings of visual discomfort or fatigue. Related studies adopted various devices or technologies and conducted surveys or experiments to acquire such data. However, those approaches are limited in this study. Therefore, this research largely used data from Metacritic, where the comments describe players' feelings about video games, and the system requirements of games from Steam.

Logistic regression is the primary method used for analysis. The model with the best performance based on the data has about 0.66 accuracy, which evaluates the ratio of correctly predicted instances to the total number of instances in the data. This value indicates that this

model does not have a perfect model prediction but is much better than random chances. The recall per class measures the ability of the model to correctly identify positive instances. The reported recall for causing discomfort is 0.45 and the recall for not causing discomfort is 0.81, meaning that this model has decent performance on classifying games that do not cause mental discomfort but not promising on classifying games that cause mental discomfort. Possible reasons causing this might be inadequate data, unreliable quantification of variables such as those nominal numbers (operating system, memory, and storage), and biased fill-in of missing values.

Based on the model, it appears that the storage and release year of games are the outstanding factors in causing mental discomfort based on the variable importance plot and their coefficients and p-values in the model summary (only these two have small p-values). This means that a game with a higher storage requirement has a higher probability (positive coefficient) of causing mental discomfort. Also, a game that was released recently has a lower probability (negative coefficient) of causing mental discomfort. Intuitively, games with large sizes are more likely to cause mental discomfort, like “Call of Duty: Vanguard”, “Call of Duty: Black Ops Cold War”, and “Call of Duty: Modern Warfare”, which are all the series of Call of Duty, and this series is typically well-known for their magnificent display and rapid scene transitions. All these games have more than 170 GB of required available hard drive space on computers. A higher requirement of available computer space means larger storage of game data. Based on my personal experience of games and the fact of game storage, games with online components and spectacular scenes are more likely to require more storage, and the game size increases as time goes by. This is because these games either need to save game data and user stats for matches with other players or need to save an extremely large number of models and

programs on local computers. On the one hand, for example, pure online games like “League of Legends” are only about 7 GB large initially, but the whole game requires much more space to save data for matches or replays over time, and these online games indeed cause players to feel uncomfortable after long-time playing because matches or cooperations with other players requires efforts and energies to put in. On the other hand, for massive games that do not have online components but still require large available spaces, they typically have magnificent scenes and require a huge number of programs to support them. Additionally, these kinds of games normally have a longer length of the campaign – the mode and story for single players. These contents are generally attractive, and players are likely to spend a long consecutive time playing them, which increases the probability of causing uncomfortable feelings and poor sleep quality (recalling the game and feeling excited while sleeping). Therefore, games with higher requirements of storage are more likely to cause mental discomfort, while the other requirements are less essential based on the analysis in this study.

It appears to be confusing that recently released games are less likely to cause mental discomfort and it also contradicts the hypothesis. Possible explanations can be that game developers have recognized the fact that realistic games are more likely to cause discomfort symptoms, so they intentionally added some elements that could reduce negative effects, like the elements used in video games for reducing vertigo (mentioned in the literature review). Additionally, with the boom of video games in recent years, the group of players that always play newly released games might be less sensitive to the magnificent effects from screens and therefore less likely to report any discomfort, while the majority of players may not try those new games unless a certain number of connoisseurs have recommended them.

Moreover, it is unexpected that graphics play an unimportant role in the models since graphics cards are always the representation of the high performance of computers, and better graphics cards are always the meaning of high-end technology. This unexpected result is potentially caused by the missing data. The missing values are intentionally filled with values that may not be the original requirements of games, and the difference between the benchmark of the latest graphics card and outdated ones is huge, which ranges from 0.05 to 148. Furthermore, the comments data might have some issues of independence. As mentioned earlier, players' comments on games can be potentially affected by others' comments, for example, due to the nature of commenting and herd mentality, if games do not have too many comments and all the comments are positive, players are likely to be affected by others' evaluations and avoid the minor defects of the games.

The logistic model with interaction terms is also attempted in the analysis. An interaction term represents the effect of the combined influence of two computer requirements on mental discomfort, which indicates that the effect of one requirement on mental discomfort depends on the value of another requirement. For example, the effect of storage requirement on mental discomfort depends on the specific brand and model of graphics cards, and this effect can be different across different brands and models of graphics cards. The model with interaction terms in this study is not perfect because none of the terms are significant. This situation can be improved if more data can be analyzed. However, solely based on the model, interaction terms such as storage and memory, graphics card and operating systems, storage and graphics card, etc. are important factors in evaluating whether games cause mental discomfort or not. The model with interaction terms tells a more specific and comprehensive story about the correlation

between mental discomfort and computer configurations, and it might become reliable with more available data.

5 – Conclusion

5.1 – Limitations

In this study, several limitations derived from assumptions throughout the research. For the comments from Metacritic, a few words are considered as the keyword that players may use to express their feelings of mental discomfort, these words include “dizzy”, “dizziness”, “headache”, “tired”, “eyesore”, and “fatigue”. It is assumed that players write these words to express their literal meaning, but the cases in which players may say “not tired” or “no headache” are ignored. They may express the opposite feelings, but they were recognized as having the same expressions. Also, these words could only be a small fraction of expressing feelings of mental discomfort, so the number of comments that express such negative feelings is likely to be underestimated.

For system requirement data from Steam, the recommended system requirements for each of the games are used for analysis. However, for some of the games, they only have the minimum system requirements, which causes the data to be missing from the collection. Some games have all the five system requirements, operating system, processor, memory, graphics card, and storage, while some games only have requirements like operating system and memory. Since the missing values were filled in manually, like “2” for missing graphics cards, some of the games may have better hardware written in the minimum system requirements section that were not scrapped, or their format of scrapped text are highly time-consuming for extracting the corresponding strings. The adopted fill-in approach may not capture the nuances and specific

requirements of different games, and there may be some limitations to using this method for more complex analyses.

Additionally, mental discomfort does not necessarily depend on the games themselves but on how players behave and the variables in the data do not tell the complete story of games. In fact, people may feel dizzy or eyesore even working with computers and looking at screens for a certain amount of time. The screen itself is responsible for such discomfort symptoms in many of the scenarios. In this study, players' behaviors like the length of gaming sessions, lack of water, posture, etc. are not considered as the hidden factors causing mental discomfort. Furthermore, other factors that may be the possible cause of mental discomfort are not included in the data, like the field of view, default frame rates and resolution, and the player's perspective (first-person perspective, third-person perspective, etc.). These variables may enrich the analysis and tell a more comprehensive story about the correlation between mental discomfort and game requirements.

5.2 – Future work

With the recognition of the limitations that are mentioned above, several aspects can be improved to enhance the credibility of this research. Firstly, the numeric values generated based on Metacritic comments could be more comprehensive if additional keywords that represent mental discomfort are considered. A more complex program could be developed to include those that express true feelings of mental discomfort to increase the accuracy of the values. Secondly, while graphics card and processor data were benchmarked from reliable online sources, quantifying the benchmarks for the operating system, memory, and storage requirements would enhance the credibility of this work. Having reliable sources to quantify these three requirements

can improve the credibility of this work, regardless of the loose requirements for operating systems, memory, and storage. Thirdly, with over 5,500 games from Metacritic and Steam, the combined data only includes approximately 1,400 games. Merging the two data sets and increasing the sample size could provide more comprehensive results.

There are also several interesting branches of this research worth pursuing. For example, 1) exploring the effect of computer configuration on specific keywords, 2) evaluating the change of negative effect over time, and 3) investigating the story behind each comment. For the first example, creating two datasets, one with the keyword “tired” and another with “dizzy”, could tell which exact mental discomfort players may have with games that have certain computer requirements. For the second example, conducting a time series analysis to see whether more games are reported to cause mental discomfort during the past decades and predicting the future trend could provide valuable insights. For the third example, collecting all the comments of each user who has commented on any of the PC games could allow for exploration of the stories behind mental discomfort. Caution should be exercised when considering and weighing comments from users who frequently mention a specific type of mental discomfort to account for any underlying factors.

6 – Acknowledgements

The authors would like to thank Prof. M. Lavin for his guidance on this research.

7 – References

Akowuah, Prince Kwaku, et al. "Digital device use, computer vision syndrome, and sleep quality among an African undergraduate population." *Advances in Public Health* 2021 (2021): 1-7.

Awati, Rahul. “What Is Resolution and How Is It Used?” *WhatIs.com*, TechTarget, 3 Aug. 2022, https://www.techtarget.com/whatis/definition/resolution.

Brunner, Doug. “Frame Rate: A Beginner's Guide.” *The TechSmith Blog*, 27 Sept. 2022, https://www.techsmith.com/blog/frame-rate-beginners-guide/.

Claypool, Mark, and Kajal Claypool. "Perspectives, frame rates and resolutions: it's all in the game." *Proceedings of the 4th International Conference on Foundations of Digital Games*. 2009.

Eskasasnanda, I. Dewa Putu. "Causes and effects of online video game playing among junior-senior high school students in Malang East Java." *Komunitas: International Journal of Indonesian Society and Culture* 9.2 (2017): 191-202.

Hoffman, David M., et al. "Vergence–accommodation conflicts hinder visual performance and cause visual fatigue." *Journal of vision* 8.3 (2008): 33-33.

Kanyılmaz, Tuba, et al. "Effectiveness of conventional versus virtual reality-based vestibular rehabilitation exercises in elderly patients with dizziness: a randomized controlled study with 6-month follow-up." *Brazilian journal of otorhinolaryngology* 88 (2022): S41-S49.

Lambooij, Marc, et al. "Visual discomfort and visual fatigue of stereoscopic displays: a review." *Journal of imaging science and technology* 53.3 (2009): 30201-1.

Metacritic. *Best PC video games of all time*. Metacritic. Retrieved April 7, 2023, from https://www.metacritic.com/browse/games/score/metascore/all/pc

Phillips, J. S., et al. "Vestibular rehabilitation using video gaming in adults with dizziness: a pilot study." *The Journal of Laryngology & Otology* 132.3 (2018): 202-206.

Scheiman, Mitchell, and Bruce Wick. *Clinical management of binocular vision: heterophoric, accommodative, and eye movement disorders*. Lippincott Williams & Wilkins, 2008.

SteamDB. *Steam charts · Most played games on steam · steamdb*. Retrieved April 7, 2023, from https://steamdb.info/charts/

Steam. *Steam store*. Welcome to Steam. Retrieved April 7, 2023, from https://store.steampowered.com/

Userbenchmark. *CPU UserBenchmarks - 1404 Processors Compared*. Retrieved April 7, 2023, from https://cpu.userbenchmark.com/

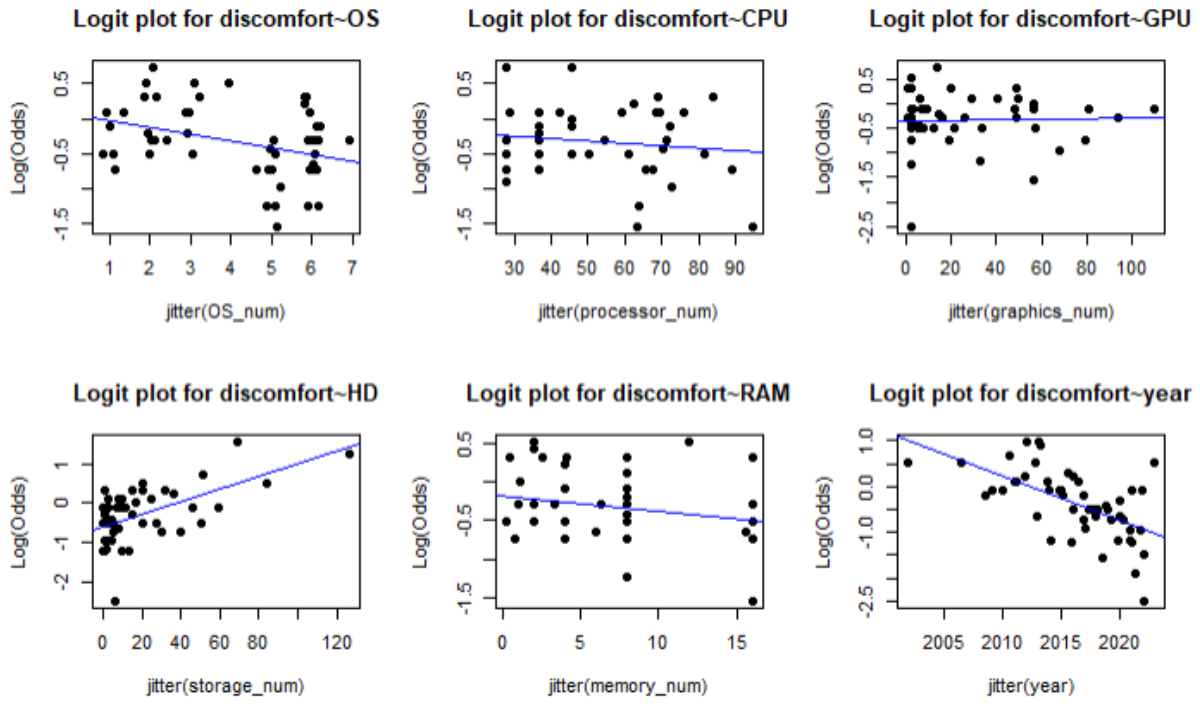
Userbenchmark. *GPU UserBenchmarks - 698 Graphics Cards Compared*. Retrieved April 7, 2023, from https://gpu.userbenchmark.com/

Viirre, Erik, and Robert Sitarz. "Vestibular rehabilitation using visual displays: preliminary study." *The Laryngoscope* 112.3 (2002): 500-503.

Zhou, Jun, et al. "Eye movements and visual discomfort when viewing stereoscopic 3D content." *Digital Signal Processing* 91 (2019): 41-53.

8 – Appendix

GitHub Repository: https://github.com/DQuan759/DA-401-First-Draft.git



Appendix Figure 1: Empirical logit plot for OS, processor, graphics, storage, memory, and release year without transformations

	True Values		
Predicted Values		0	1
	0	212	134
	1	20	35

Appendix Table 1: Confusion Matrix for sys_req_mod

	True Values		
Predicted Values		0	1
	0	186	102
	1	45	64

Appendix Table 2: Confusion Matrix for year_mod

	True Values		
Predicted Values		0	1
	0	187	91
	1	44	75

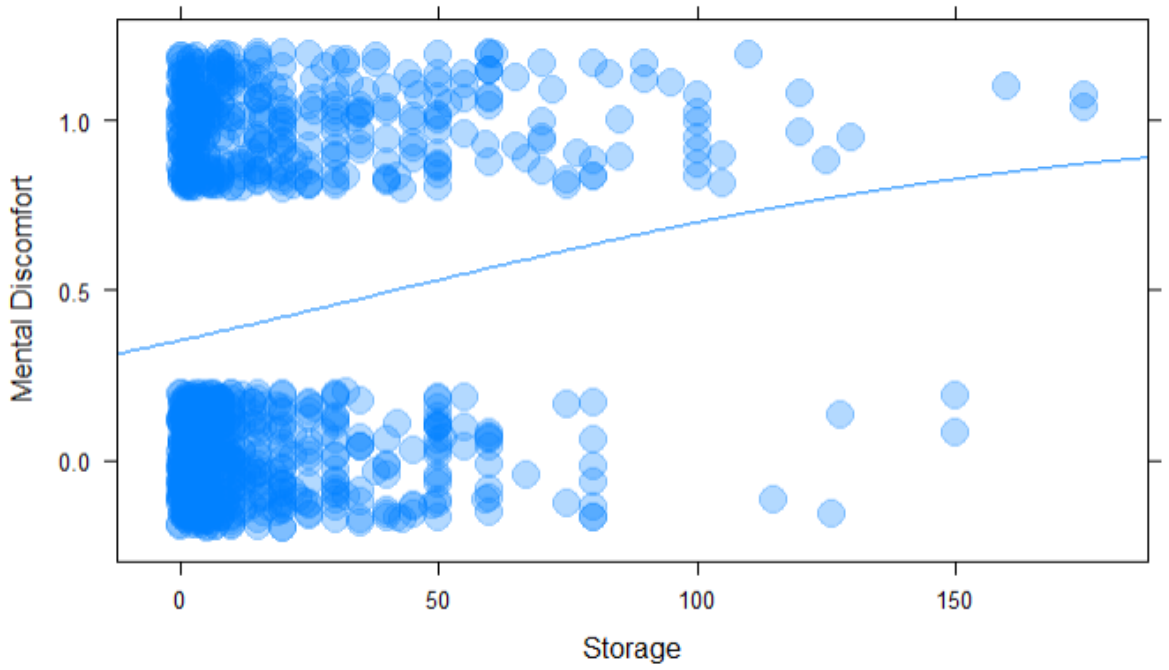
Appendix Table 3: Confusion Matrix for trans_mod

	True Values		
Predicted Values		0	1
	0	183	85
	1	48	81

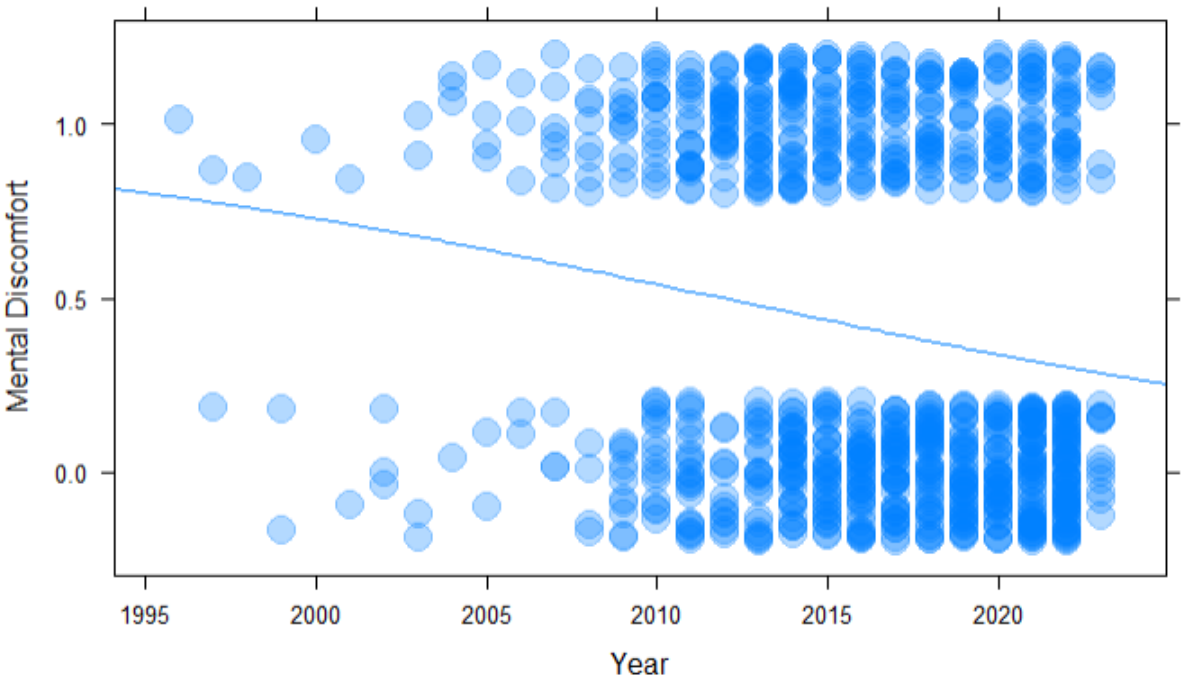
Appendix Table 4: Confusion Matrix for interact_mod

	True Values		
Predicted Values		0	1
	0	181	82
	1	50	84

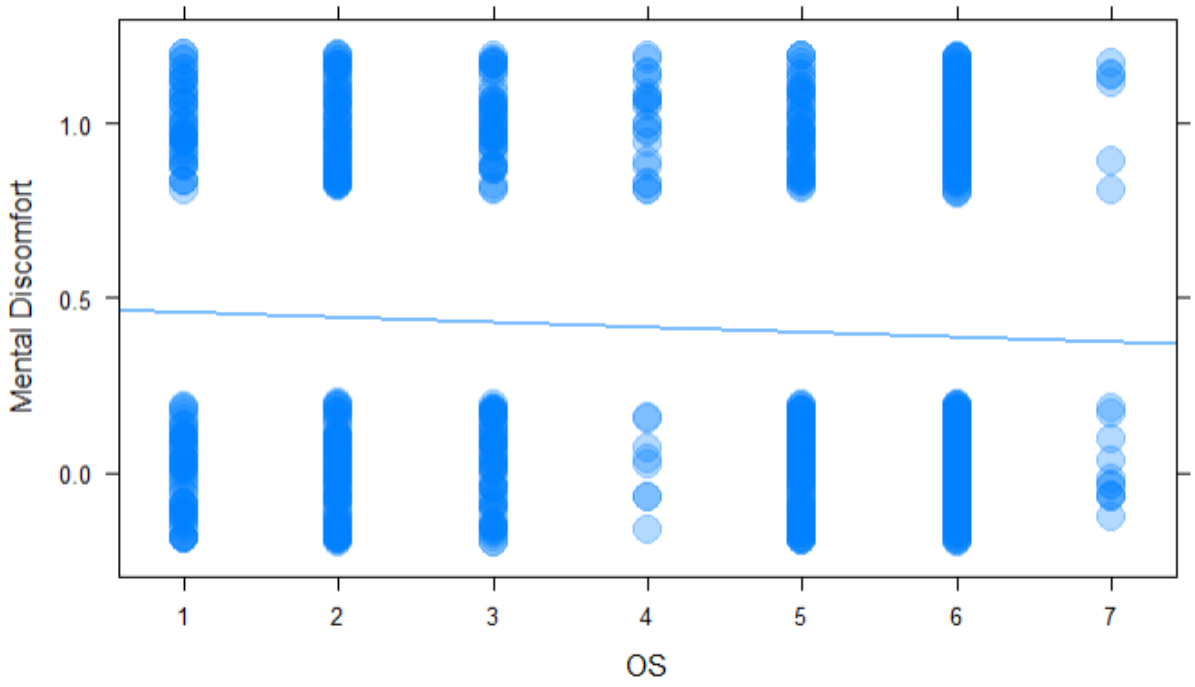
Appendix Table 5: Confusion Matrix for both_mod



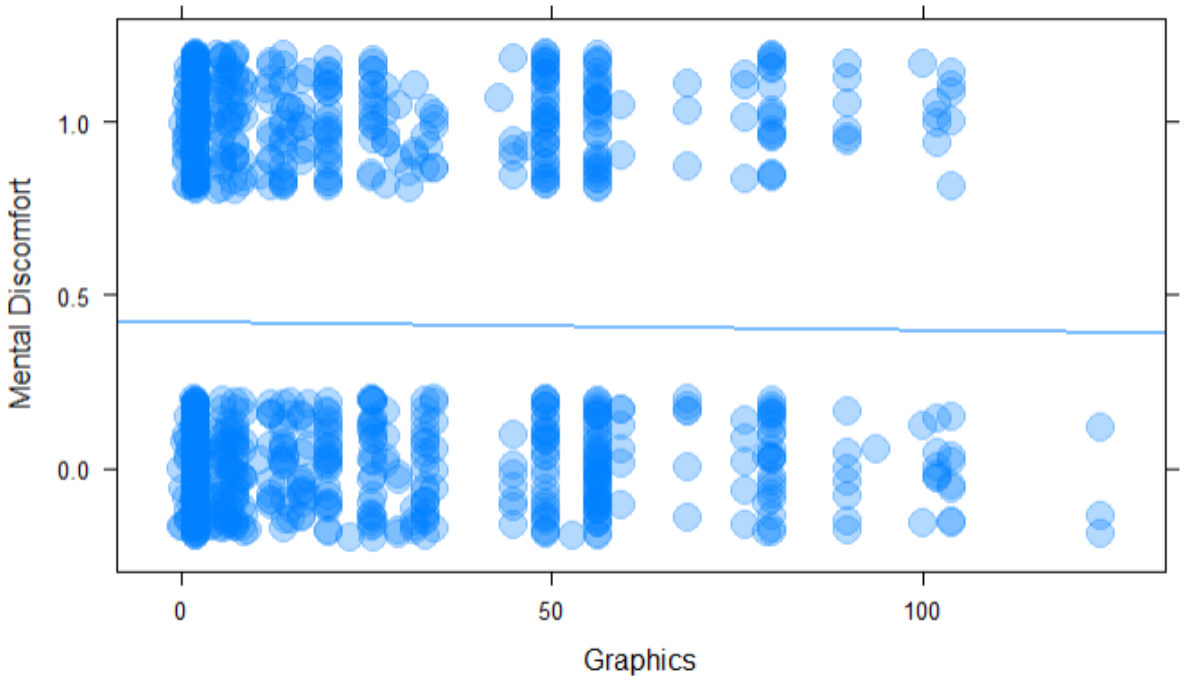
Appendix Figure 2: Scatter plot of Storage with overlaid logistic regression line



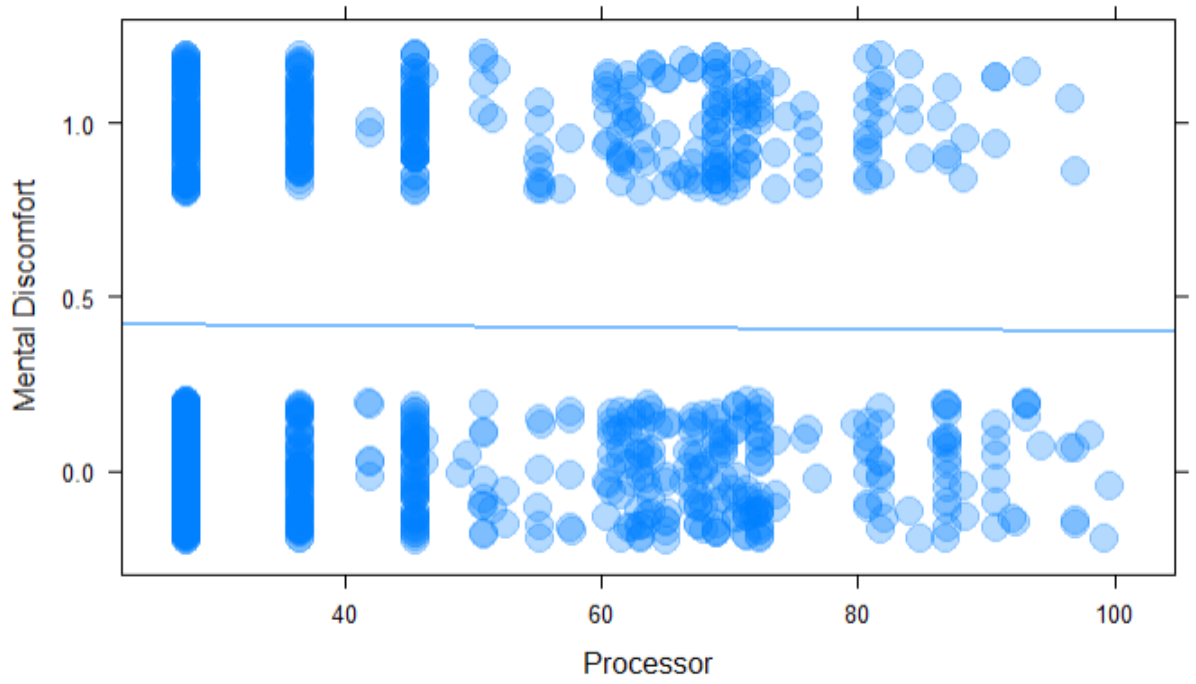
Appendix Figure 3: Scatter plot of Year with overlaid logistic regression line



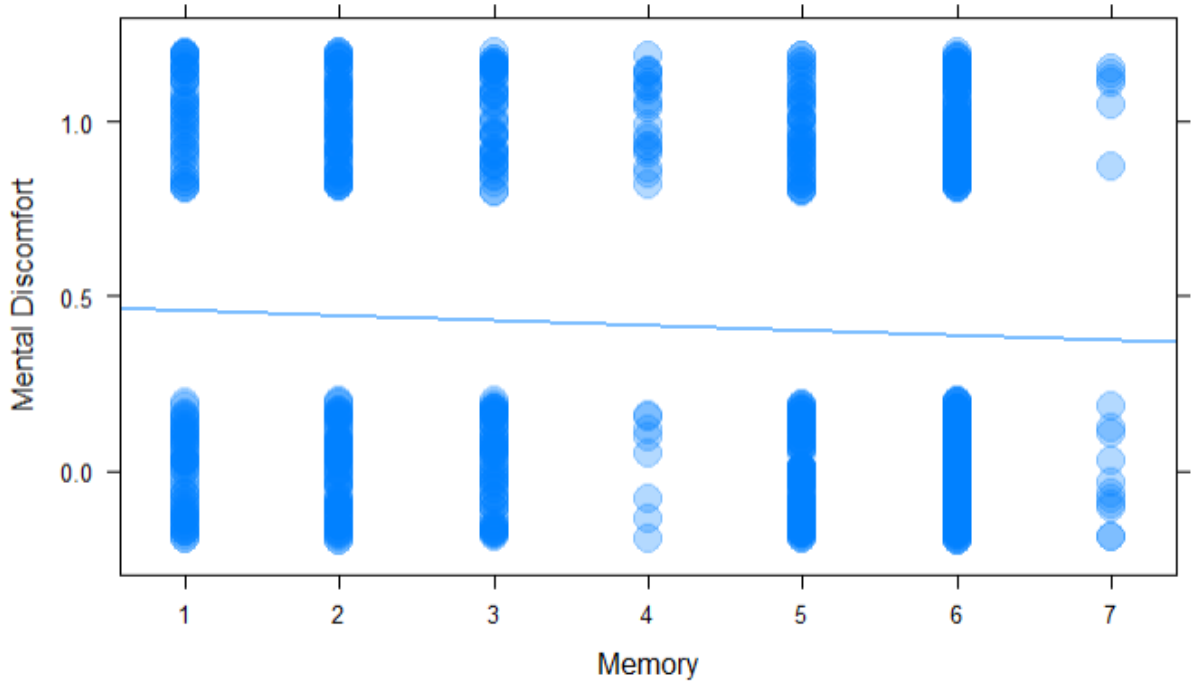
Appendix Figure 4: Scatter plot of OS with overlaid logistic regression line



Appendix Figure 5: Scatter plot of Graphics card with overlaid logistic regression line



Appendix Figure 6: Scatter plot of Processor with overlaid logistic regression line



Appendix Figure 7: Scatter plot of Memory with overlaid logistic regression line