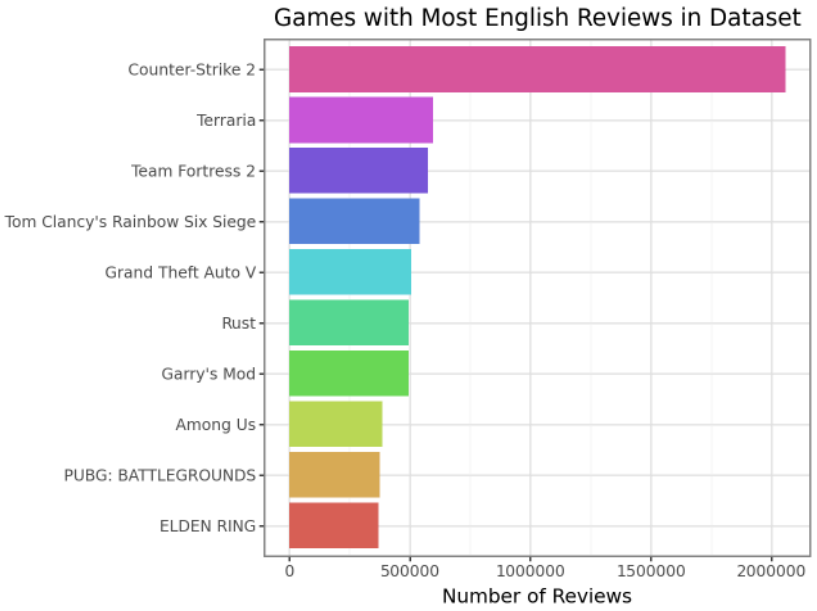
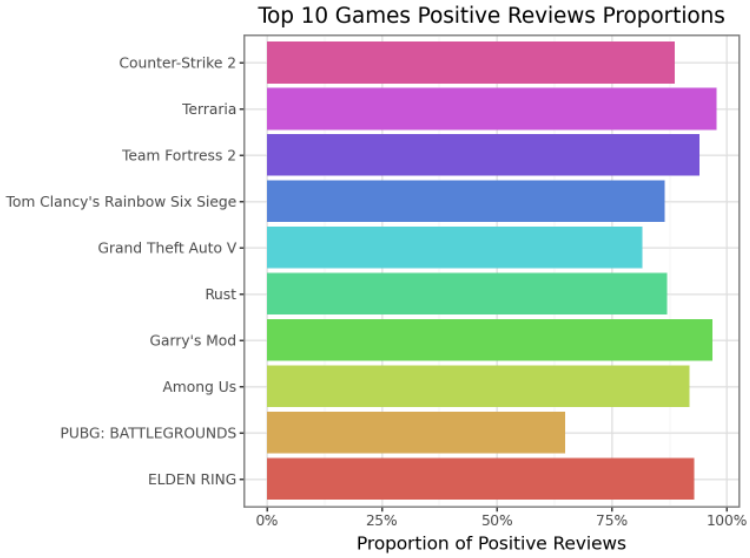
An Analysis of Toxicity of Reviews on Steam Games

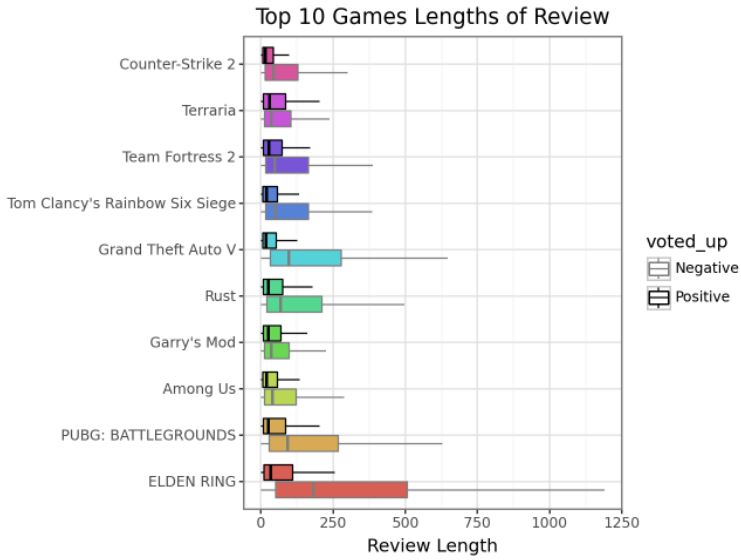
As a student who's so close to FINALLY having time off during Spring break, I really want to start playing a new game. When it comes to game selection, I'm the type of person that obsesses over new hobbies, spending hours upon hours immersing myself, and plunging myself into a new community where I can talk with others about my new hobbies. Due to this, I want to get into a game that doesn't have a toxic fan base, I've already been burned too many times with games like League of Legends. So, using this dataset, I'm going to analyze and assess the toxicity within video games just based on their reviews.

To familiarize myself with this dataset, I ran some exploratory analysis. I observed that Counter-Strike 2 had a massive representation in the most popular games, with other popular classics such as Terraria and Team Fortress having a lot of reviews as well. For simplicity, I want to focus my analysis in game selection to this top ten, as I want a game with a large and thriving community.

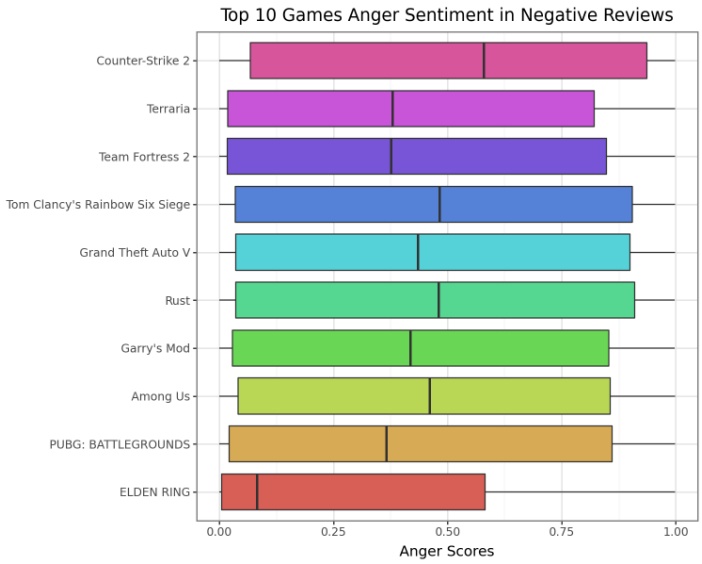


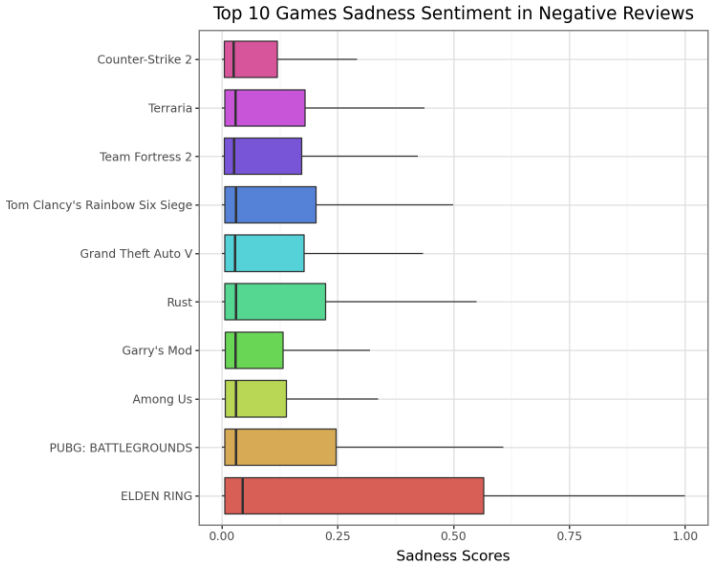
Afterwards, I wanted to check on some heuristics. I noticed that PUBG had a significantly lower positive review proportion than any other game in the top ten. I'm not too sure why, but I do remember a lot of buzz of a cheating scandal existing within the game that made the player base full of hackers that could insta kill any player given the right weapons. Additionally, whilst every game had relatively similar positive review length, I noticed that Elden Ring had significantly larger negative reviews.





However, despite these findings being really insightful, these were all just based on heuristics, not the reviews in context. To determine contextual evidence, I decided to run a sentiment analysis that would be able to assign scores of happiness, sadness, anger, fear, shock, and love. I limited my sentiment analysis to negative reviews, as that would capture the most toxic fans of a playerbase. The two emotions that I wanted to focus on the most were anger and sadness, as anger would definitely be a proxy towards toxicity, and sadness could be considered a reverse.





As shown, it was very clear that Elden Ring had the least angry negative reviews, as well as being the most sad. It was also apparent that CS2 had the most angry reviews. This makes me believe that Elden Ring might be the game of my choice to play over spring break.

The last thing I wanted to focus on is how the emotions of these reviews are associated with playtime length, and how it specifically affected my choice of game, Elden Ring. Just from previous knowledge, I know Elden Ring is a very difficult game, so I wanted to see if players being angry was more so associated with the fact that they thought the game was too hard and quit too early. I decided to run a negative binomial linear regression to assess this. To interpret a model such as this, we have to view the sign of the coefficients to determine whether it's associated positively or negatively, as well as the p-values to determine if there's statistical evidence to suggest association at all.

Emotion	Coefficient	P-value
Sadness	-.3007	.001
Joy	-.3442	<.001
Fear	-.6916	<.001
Surprise	-.7333	<.001
Anger	-.0864	.362

With all except Anger having p-values of <0.01, there is very strong evidence to suggest that sadness, joy, fear and surprise emotions are all associated with less play time. This goes against my initial assumption that people who leave angry reviews were just mad that the game was really difficult and quit too early. In actuality, the association exists with joyous, sad, and fearful negative reviews