Joe Lollo

LIS 572

Professor Melanie Walsh

14 November 2022

Dataset Biography

Dataset Source: https://github.com/ChessPiece21/LIS-572/blob/main/Final/ow2-2022.csv

This dataset comes from Overbuff, a fan site featuring rich statistical analyses of Activision Blizzard's *Overwatch* game including players' win rates, average rankings, and average performance using different player characters (referred to as "heroes"). The data was collected by site administrators for Overbuff's public audience – it was created by *Overwatch* fans for *Overwatch* fans – it came from game-wide public records of player pick and win rates across different platforms, although no specific distinction for Xbox, PlayStation, PC, or Switch players is made in the data, and was collected by volunteer moderators and administrators for the site. This site treats *Overwatch* and the fan culture surrounding it as a "sport" of sorts, and as a result the database features some very in-depth data tables similar to those found on sports fan sites, but I went with the general one that included a mix of qualitative and quantitative variables. The data was collected so that *Overwatch* players could gain a better understanding of how their community played, especially in more competitive settings. The dataset is the raw CSV version of the data featured in Overbuff's home page, but it is featured on my GitHub page for reasons you will read in the next paragraph.

While the dataset is public-facing at the moment, I had to request the data as a CSV from the site's administrators, through a link at the bottom of the dataset, rather than downloading a CSV directly from the site's statistics page because of Overbuff's policy regarding the usage of the data – I had to explain in an email that I was a graduate student interested in doing data analysis on the site's documented *Overwatch* hero statistics, included some of the questions I had from the last assignment, and left a disclaimer that I was going to

feature both the CSV and all of the analysis on my GitHub page. This led to one of the site administrators approving me and sending the CSV file over, as long as I sent the GitHub link in a follow-up email once I uploaded it. Something I find very interesting is that this CSV file has a few columns that are not posted on the public-facing website, including each character's gender and average difficulty rating, the latter of which is indicated on the *Overwatch* website. The main dataset is rather limited in its public-facing form, but the addition of more categorical variables in the CSV I received makes it even better because there are more columns, and therefore more questions to consider. My analysis will take a "narratives beyond the numbers" approach as discussed in *Data Feminism*, as I will attempt to interpret *why* players will pick certain characters and denote any biases among the quantitative data points, as each of these rows and columns can tell stories about how *Overwatch* players engage with the product.

I manipulated the data through a Google Sheet in one way, and it only came after I did some exploratory analysis of the data in RStudio – because I saw a problem. Specifically, there were two characters' names, specifically D.va and Soldier '76, revealed errors in the data analysis when trying to manipulate the data using R functions, something I was partially expecting because of the way that these characters' names are written in-game. I toyed with a few different ideas, but ultimately came to what you will see in the GitHub link – dashes where the dot and apostrophe are. I thought that cleaning the data was the best option I had available, because then I could be able to analyze the entire CSV data frame, even if the names aren't 100% accurate to the characters' names at the moment. All other data manipulation I see myself doing using the DPLYR functions. There is an additional limitation that comes with the short time span of the dataset – since *Overwatch 2* has been out since early October 2022, there has been only two months of data collection and analysis, updated on a weekly basis – I don't have the final version of this dataset, for obvious reasons, yet I think this gives a good snapshot of what the recent state of esports is like, despite the small sample size and small amount of recorded data.

An ethical complication in this dataset comes from Overbuff's history with Blizzard, the developer of *Overwatch*, especially the new relationship they formed – while Blizzard sent cease and desist letters to Overbuff as well as other fan sites and creators in 2020, citing "improper" usage of their copyrighted content, as of November 2022, they seemed to have had a change of heart. Blizzard developers began collaborating with the fan site to turn it into a hub for *Overwatch*'s esports community, which has been rising since the release of *Overwatch 2* in October. Due to the tumultuous relationship between the two, and Blizzard's infamy for taking down many fan projects that use characters from its properties, even if it's for something purely creative and shareable among fan communities, I have to be wary with what exactly I include within my analysis and where I post the final results, since I am wary of the potential copyright strikes that may arise.

Overall, being an *Overwatch* fan who's involved in the game's fan community in various ways, I am excited to begin exploring this dataset, and especially excited to use this data for good and come to a better understanding of the many factors that make up today's *Overwatch* scene.