Derek Chen
Chetan Gupta
John Proserpio

CS 3300
Project 1 Write Up

Amazon 4th Gen Console Reviews Visualization

A. The data that we utilized to create our visualization is Amazon product reviews data for 4th generation consoles. We scraped review data using Node.js and a Node module known as jsdom, which allowed us to retrieve relevant information using the dom structure of the page. In doing so, we could grab information about a review, such as the text, the author name, the date, the rating, the helpfulness etc. We used this in the included file AmazonScraper.js to scrape reviews into a JSON file. We had to do this multiple times as the scraper was prone to failing sometimes due to errors from the Amazon server. Eventually, we scraped a total of 6,000 Playstation 4 reviews, approximately 3,600 Xbox One reviews, and approximately 550 Wii U reviews. We had to scrape two separate editions of the Xbox One as one version was the current edition being sold, while another was a special “Day One” edition that was a launch version sold when it first came out. We combined both review sets for the Xbox One to get a new review set of approximately 3,600 reviews, as we knew they were unique. The Day One Edition only had reviews for November 2013 to December 2013. The current version had a small patch of reviews from November 2013, but the majority of it was from recent reviews the past few months in 2014. The PS4 data was all from one product as was the Wii U data.

When we scraped the reviews, we were concerned with three variables. The month that the review occurred, the year it occurred, and the actual review rating. Through gathering this information in our JSON file, we were able to load in the individual JSON files and transform it into the final data we used in data.json. The wii.json, xboxone.json, and ps4.json contain the originally scraped review data. Using these individual data sets, we loaded all the reviews into an associative array, using the first of the month that the review occurred as a key, and we aggregated the review ratings for each month while also keeping track of the number of reviews we’d summed this over. Then, we divided the total sum by the review count, to get an average review rating for each month that the product had been selling/reviewed. This data we then outputted to our final data.json file which was loaded and used to create the visual elements as described below.

B. The variables mapped to visual elements were: Time, Average Rating, and Number of Reviews. For every month, the average rating from 1 to 5 was calculated from all of the reviews scraped for that month. This number is mapped as height on the Y-axis of the graph.

Time is shown horizontally along the X-axis. We mapped each data point to the 1st of the month for which that data point existed. The time scales we used were linear and represented different timescales for the PS4 and the Xbox One than the one used for the Wii U. This was due to the Wii U’s release a year prior to the other two products.

We chose circles because we thought the radius of a circle made it an interesting shape choice. It gave us another visual element we could map a data property to. In this case, the radius of the circle. Each month corresponds to a circle on the graph, whose radius represents the number of reviews that were submitted for that particular product in that month; a bigger circle means more reviews. We set the minimum radius to be a radius of 6, so as to ensure even months with sparse review counts would still be visible. We figured the scale that the larger circles used, along with our legend, helped show that even the smallest increase in circle radius had a huge impact on the review density it represented.

To help reinforce the positivity/negativity of reviews, a gradient color scale was implemented in which a 1-star average rating creates a red data point, and green for a 5-star average. 3 stars correspond to yellow. This gradient goes along with the convention that people perceive green as good and red as bad, while yellow is looked at as a neutral color. Reviews in between 1 and 3 would be an orange color while reviews between 3 and 5 were a yellow-green.

C. This visualization allows a user to see Amazon product reviews in a whole different way. Products are always changing. Instead of just seeing an aggregate bar graph for all reviews ever submitted, our illustration allows the user to see how customer opinion has changed over time. It paints a picture of the product's rating across its entire lifetime on Amazon. When some items are first released, they may be buggy or have other issues, but as time passes and people find fixes for these problems, the rating should theoretically increase assuming the core product is good. On the other hand, a product could be released with a lot of hype and get many good reviews at first, which then decline as users discover the underlying issues and bugs that plague the product.

We wanted our visualization to allow a user to be able to compare similar products and based on the average rating over time, make an educated decision on which product to buy. The goal was to make this visualization simple and easy to understand, while also being informative enough that one can make an informed decision after glancing at several charts.

We chose to visualize the ratings for three recently released competing products: the Xbox One, the PlayStation 4, and the Wii U. We thought these would be interesting products to visualize for several reasons.

1. First, video game consoles always build up a ton of hype. People get extremely excited about the next generation of their favorite console and love comparing how it is doing with it's competitors.
2. Second, the initial opinion of consoles is often vastly different from that of subsequent months. For example, the Xbox 360 came out and many people thought it was amazing. Then, everyone discovered the prevalent hardware issues known as the red ring of death. Consumers' opinion of the system plummeted. But, as Microsoft released subsequent versions of the console with improvements, it re-gained traction in the market as consumers faith in the integrity of the system was restored.
3. This is a situation we could easily find ourselves in, and therefore we can understand what the consumer would want to see and what would help the consumer make an educated decision.

When looking at the data, it is important to keep in mind that all three systems were released in

November, however the Wii U was released a year before the Xbox One and the Playstation 4. What this means is quite surprising, as the Wii U in total has somewhere around 550ish reviews for it's lifetime of 16ish months. Meanwhile, the PS4 in it's first month alone had almost 4,000 reviews! At the same time the Xbox One also had almost 1,700 reviews! Now, while reviews aren't directly indicative of product sales, one can infer that with more sales comes more reviews, so from this alone we can see that in terms of sales, the PS4 is selling much more than the Wii U and the Xbox One.

As time goes by, we notice reviews decrease each month, showing a slowing down in hype for the products, so we can now guess that these few months of data will have more accurate estimates of the actual product quality. The Wii U maintains a high quality rating for the majority of it's lifetime, however it's review count is significantly less for each of those averages. The fact that the Xbox One and the PS4 have average review values that are also quite high while still having such a larger review count, could imply a greater amount of consistency in customer satisfaction. The fact that they have so many more reviews also shows how much of a difference there is in terms of popularity of the products.

Finally, we notice that there is a slight shift for the last data point for PS4 and Xbox One treading in the mid 3s. This can be attributed to the fact that we mined data for March so the few reviews that exist for this month, are heavily skewing the average to be closer to 3 than 4. The Wii U does not have any reviews for this month, so we don't have any average review value for March for it.

Our final graph highlights the raw average review values for the three products, allowing one to compare them directly without taking review density into account. The results allow one to see the small difference in the PS4 and Xbox One averages. The PS4 not only has more reviews than the Xbox One, but it has a higher satisfaction on average as well. Meanwhile the Wii U while not nearly as popular as either device, boasts a high satisfaction rate amongst its population of consumers as well.

Based on this, if one was trying to decide between all three systems they could make a few rationalizations for their purchase decision. They might decide, based on the Wii U's high ratings, that it would make a good purchase choice given it's lower price and higher satisfaction compared to the other two systems. At the same time, one might decide it's lower popularity means it isn't the "in" system and therefore they'd narrow it down to between the PS4 and the Xbox One. At this point if they go based off the data, the consumer would likely choose the PS4 for it's greater density in reviews/popularity as well as it's higher satisfaction.

It was a bit surprising to realize how much more reviews exist for the PS4 than the Xbox One. It was also startling to realize how behind the Wii U is despite it's head start. Granted, we can't take these raw review counts and use them to estimate unit sales for the system, given there are many other retail sellers and online stores that sell these items. However, given how popular Amazon is, it does provide us a snapshot image of the position of each of these products in the race for market share, because more reviews is indicative of more popularity and likely of more purchases of a product.