

Visualizing Amazon Bestseller Data
Final Project Report – INFO 4602

Introduction

Sales figures are some of the most important sets of data in any retail business. This is because investing in sales data analysis could lead to trend insights, future sales, and increased revenue. In this project, we visualize sales data for the largest online retailers in the world, Amazon.com. With D3 visualization, we reveal pertinent trends to Amazon's business, such as their best selling products, best selling categories, and what kinds of products are predicted to generate the most revenue based on their sales to review ratio.

Data

Our data consists of cross-category monthly sales data extracted from Amazon.com. Typical sales data is unavailable to the public, but by using the Amazon API and an Amazon product research tool called [JungleScout](#), we've scraped sales data from [Amazon's top product categories](#). The top 20 listings in each product category were compiled into a master dataset and analyzed in this project. View the data under 'src/data/master.csv'.

Description

Scatterplot

The scatterplot visualization gives us insight into which products have the greatest amount of sales with the least amount of reviews. All of Amazon's top selling products from our master dataset are represented here. Upon mousing over a plot, a tooltip gives detailed information on the product, including its category, title, revenue, sales, and number of reviews. Low reviews and high sales is indicative of a niche product/market fit that's resonating with customers. A vender might consider selling

Conner Simmering
Tyler Rostenbach

products with a high sale to review ratio in order to maximize SEO and profits on Amazon.

Barchart

The barchart visualization allows a user to view the top selling product from each category based on unit sales on Amazon. Bestsellers don't necessarily mean best revenue as a product can sell many units but yield lower profits or have a higher sales margin. Revenue numbers give us more insight into the rank of a product. This visualization also allows us to easily compare bestsellers in each category by examining respective top seller revenue, sales, and reviews with a mouse-over tooltip.

Discussion

One of the questions we were trying to answer with our visualization analysis is what are the absolute best selling products on Amazon across all categories? While this is certain to change over a time, our scatterplot reveals as of May 7, 2017, this [this book](#) is the current 2nd bestseller on Amazon in terms of volume of product moved with the video games category omitted from the dataset. Various reusable beauty, office products, and other trail close behind. Moving product does not necessarily mean making the biggest ROI however. Amazon's Echo and Amazon Dot happen to be in the middle of the road for number of reviews and sales numbers (see isolated blue dots near 55k reviews), but make almost triple the revenue of the highest volume mover on the scatterplot. We hypothesis that this is because Amazon is able to conduct a fierce marketing campaign for their own products on Amazon.com.

In the barchart, we see that the video games category blows away the monthly sales numbers and revenue of all other categories several times over. This category and its products were omitted from the scatterplot in order to normalize the data, however, video games are clearly the top sellers on Amazon.