

Video Game Sales vs Scores

CS 480x Data Visualization – Final Project

Aguiar, Matthew C.

March 2, 2022

Table of Contents

1 Overview and Motivation	3
2 Data Collection	3
3 Exploratory Data and Design Evolution.....	4
4 Implementation	5
5 Evaluation	6

Overview and Motivation:

After graduation, I plan on pursuing a career in Technical Video Game Development. In fact, one day I would like to open my own game studio. However, one question that I ask myself is, do good games necessarily equal good sales? Certainly, making games shouldn't be exclusively about money, but let's face it, we need money to keep our companies running, employee's lives going as well as our own life. And with so many mediocre games these days, possibly due to companies riding on reputation and market oversaturation, I often wonder if you **need** to make a "good game"? These questions are the motivation behind this project which will aim to analyze the score games receive vs how many copies they sell. Specifically, I plan on building a data visualization to compare the critic and user scores of over 17,000 video games from 1985 – 2020, with how many copies those games sold in the U.S, Europe, Japan and combined globally.

Data Collection:

Collecting a variety of information (rating, sales, year of release, etc.) on thousands of games is not feasible in the time frame for this project so I turned to kaggle.com, a data science website which hosts datasets. The dataset can be found [here](#). It contains the following information on over 17,000 video games: Name, Platform, Year of Release, Genre, Publisher, North American Players, European Players, Japanese Players, Global Players, Critic Score, Critic Count, User Score, User Count, Developer and Rating.

Exploratory Data and Design Evolution:

Looking at the data, the best visualization to make would be a scatter plot. There are just far too many data points to make any other kind of chart. A scatter plot would allow me to see thousands of little data points and how they trend. Below is an image of the initial diagram I drew. This is something that I will be implementing in HTML, CSS, and JavaScript.

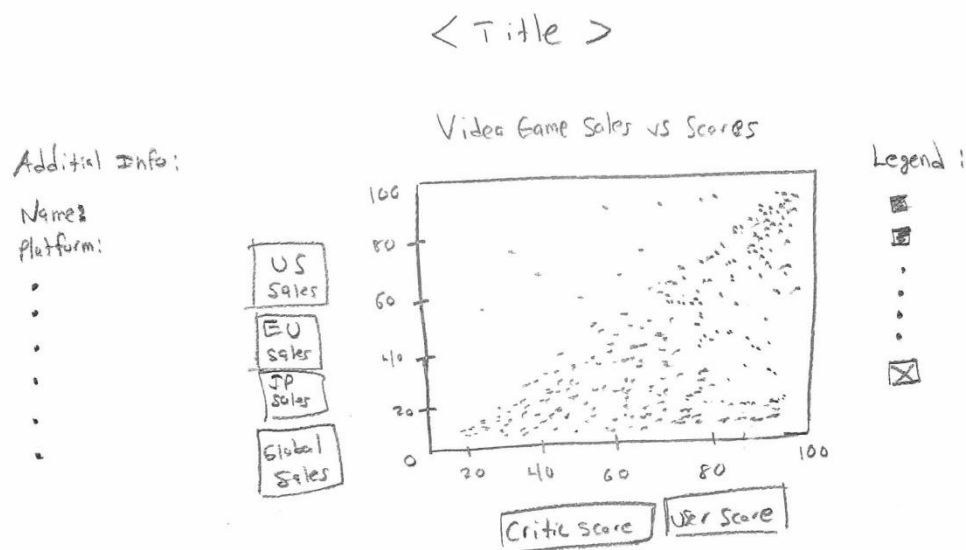


Figure 1: Planning Document for the Data Visualization UI

My initial hypothesis on how the data will trend is that higher scored games will correlate with higher overall sales. However, I think the trend will not be as strong as many think it would be. I think there will be a fair number of average games that sell well and even some good ones that

don't sell well. I think the user score will correlate more strongly to good sales in all regions than the critic score.

Implementation:

The data was visualized using HTML, CSS and the D3 JavaScript library. The link to my visualization is [here](#). It is a scatter plot where each data point represents a game and is plotted according to the game's score vs sales. The first design aspect of my visualization was to make a color-coded scatter plot. The colors correspond to different genres as shown in the legend. I thought it would be interesting to see not only do "good games" sell well, but also which genres sell well. Another design element you can see is the plus-minus button allowing the user to scale the y-axis up and down. Because there are a few outliers, the default y scale shows many of the games clumped together at the bottom, so a y-scale adjuster is a nice tool to zoom in on those smaller games. As stated in the motivation section, you can compare critic scores and users scores as well as how they compare with sales in the U.S, Europe, Japan and global sales. The UI allows the user to combine these various aspects with each other by clicking on buttons on the y-axis to change regional sales and on the x-axis for critic and user scores. Finally, the main interactive element of the visualization is the ability to hover over each data point and a side panel of data is generated showing more details about the particular game. This way users can look for even more trends beyond simply scores vs sales.

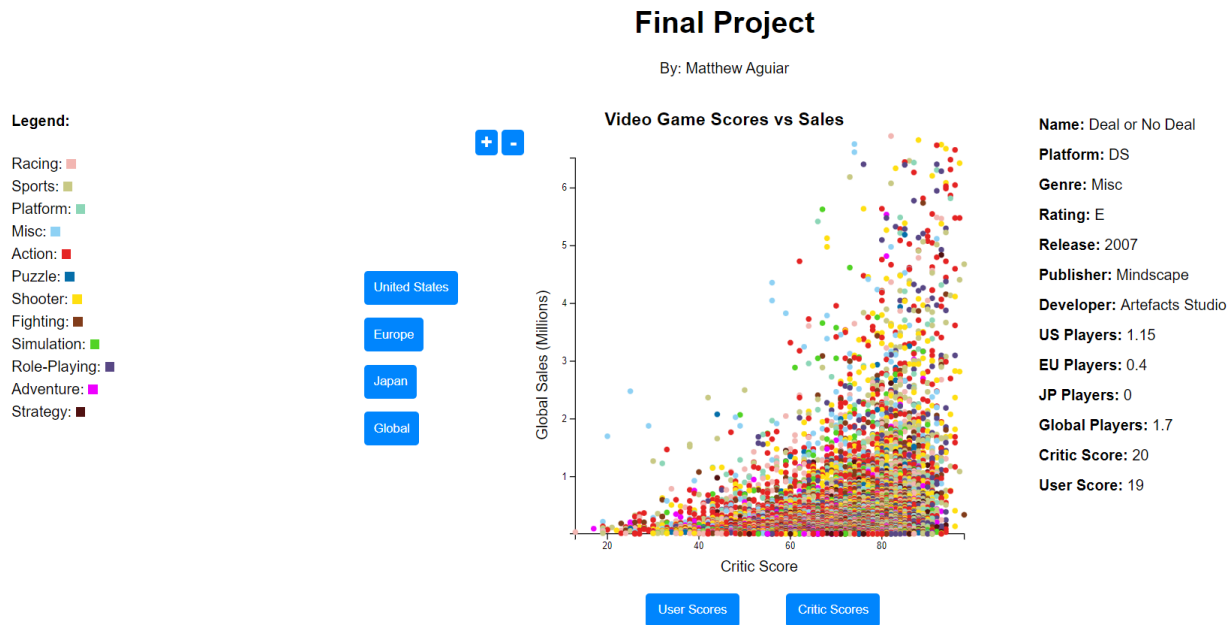


Figure 2: The final scatter plot of games

Evaluation:

After analyzing the graph my hypothesis does appear to hold as there is clearly a trend of higher scored games with higher sales. This holds for all regions and for both critic and user scores. However there does seem to be many marginal games that still sell millions of copies. One observation from hovering over the data points is that a lot of mediocre games are family-oriented games such as Wii games like “Michael Jackson: The Experience” which was rated 56 by critics and sold 4.3 million units globally and “Carnival Games” rated also at 56 by critics and sold 4.05 million units globally. This is actually something important for me to keep in mind as I would eventually like to develop family games for all to enjoy. I am not known to be a hardcore shooter gamer. Another trend is that some games sell well even with a poor score when the game is about a popular TV show or game show. For example, “Deal or No Deal” was released on the

Nintendo DS in 2007 with a critic rating of 20 and a user score of 19 but it still sold 1.7 million units worldwide.

After analyzing the data, I've learned that it is important to make a "good game". You can't throw something together with no effort. However, it is possible to make a game that sells well without it being a masterpiece. It seems that a lot of family games sell very well regardless of rating. Looking back on this project there were lots more questions I wanted to answer but not enough time. Specifically, there were lots of time related trends to see how gaming trends have changed over time such as: most popular game by year, most popular genre by year, and comparing game scores over time. Luckily, this project seems to be a good base to start with and these are questions I can adapt my website to answer in the future!