

INFO 3300 - Project 2 Writeup

Daniel Rusk
Cornell University
Ithaca, NY
dgr73@cornell.edu

Kevin Sun
Cornell University
Ithaca, NY
ks933@cornell.edu

Sarah Shin
Cornell University
Ithaca, NY
sjs457@cornell.edu

Valeria Marqués
Cornell University
Ithaca, NY
vmm49@cornell.edu

THE STORY

The story of our visualization shows us the relationship between game sales with console and genre by year. We are exploring the question: How does the type of console or genre affect the sales of games? For example, we were interested in seeing how action game sales can be impacted by the console/platform it is on. We hypothesized that certain genres, such as platformers, would have higher sales within Nintendo's consoles (Wii, 3DS, etc).

Thus, we decided to present three graphs to provide a comprehensive view of this story. The first graph, the scatterplot, allows the user to see the relationship between genre and copies sold for the first 1,000 top ranked games. Secondly, the bar graph shows how much each console had dominated the industry sales for each year from 1980 to 2015, thus showing the relationship between console and copies of games sold. Lastly, the line graph combines the two stories and allows the user to see the relationship between copies sold, genre, and console through interactivity.

What we have found through our visualization was an increase in action and sports sales over the course of the 1980 to 2015 time period, as well as the decline in popularity for the adventure games. Furthermore, we also found that, while Nintendo largely dominated the video game industry from 1985 to 1995, they had a much decreased presence from then on, with Sony quickly gaining market share. Within games on Sony's consoles, we found that the most popular genre was action, leading us to speculate on the possibility that the PlayStation console line was particularly ideal for such action games.

DATA

The data was sourced from Kaggle [[Source: Kaggle](#)]. It consists of sales and metadata about games that had sold more than 100,000 copies as of mid 2016. Variables include the game's name, overall rank, platform of release, year, genre, publisher, and unit sales in North America, Europe, Japan, the rest of the world, and globally. Overall rank was determined from global sales. Each game was attributed to a single largest platform, even if the game was released on multiple consoles.

Although the dataset was created in 2016, there existed data points up to 2020. Upon analysis, the games in these later years were misattributed and had actually been released many years before. In response, all data points from after

2017 were filtered out. Data points from 2016 were also removed because the information was not up-to-date enough to provide a complete-enough picture of the games released that year. Additionally, entries which did not provide a year of release were also removed from the data set.

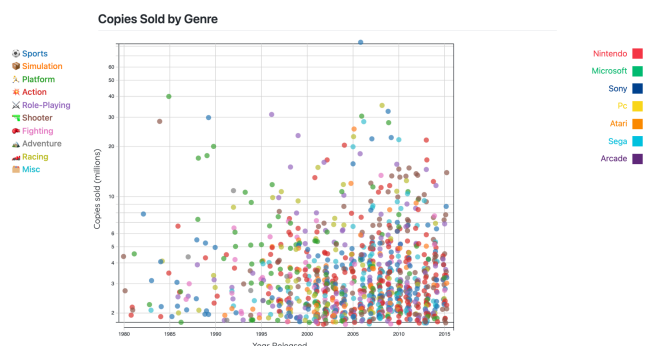
Data was also removed for our visualization purposes. Three companies: Bandai, NEC and 3DO created consoles whose games, while popular enough to enter the dataset, failed to make any visible impact on the market. In addition, strategy games made very little in sales compared to the other genres. Our visualization is much more interested in the larger picture of the video game market. Therefore, these entries were filtered from the data set.

STATIC DESIGN RATIONALE

Based on our data, we designed our graphs around answering the single question of how two factors, genre and console, can affect game sales over a period of time. With our broad dataset, we could focus our audience's attention on the several key aspects we wished to highlight through two narrower graphs, while having one overall graph that shows the compilation of all the different aspects we wished to portray. To serve this purpose, we designed three different graphs: a scatterplot, a stacked bar graph, and a line graph.

One intentional overall design choice made for our visualization was the colors. The colors we ended up using for all our graphs were standard and did not have biases associated with them, which is why we chose them for our graph depicting video games.

SCATTERPLOT

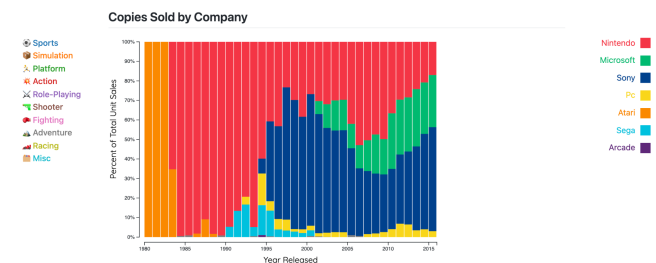


The scatter plot depicts the story of sales by genre for the top 1000 games in terms of copies sold from the years 1980-2015. After filtering, the dataset still contained over 16,000 entries. While this was fine for some of our graphs like the bar graph which aggregated all the data, this was simply too many entries to make an effective visualization for the scatterplot. In response, the top 1,000 games were used to make this graph. We feel like it still provides enough information to tell the story of video game genres over time without overwhelming the viewer and crowding out the relevant information.

The marks of the graph are the circles, or scatter plot points. The channels are the x and y position of the points that change based on the number of copies sold for the particular game and the year in which the game was released, respectively. A log scale was used on the y-axis to better convey how much a video game sold compared to its peers and prevent a few outliers skewing the entire visual. The color of the point corresponds to the game's genre. This design highlights not only the increase in the popularity of video games over time, but also how the popularity of certain genres rises and falls. The density channel incorporated into the points further highlights popular genres.

The legend, located on the left, lists all the genres present in the graph to tell the user what kind of games they can expect to find. The color of the text is the same as the color of the genre's circle - thereby making it easier for the viewer to pick out a specific genre from the graph. Emojis were also added to the legend to differentiate the genres from each other, particularly for those who may not have as much familiarity with video games.

STACKED BAR GRAPH



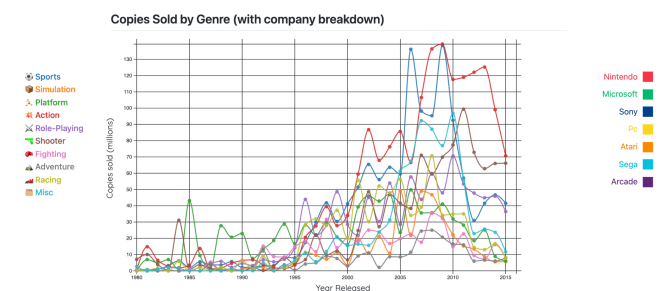
The stacked bar graph depicts the story of market share held by companies from 1985 to 2015. The marks are the bars while channels include length and position on the x-axis. The higher the percentage a company had of a year's video game sales, the longer its bar becomes. This reinforces the idea that the larger the company's sales, the larger it's bar is. The x-axis follows time, thereby making a visual narrative as certain companies rise and fall in popularity according to the size of their bar on the graph.

The color of the bar is also a channel, corresponding to an individual company and all of its related consoles. For

example, everything from the NES to the WiiU is contained under Nintendo. This provides consistency through console generations and prevents having to identify each individual console (of which there are over 30). The colors were specifically chosen to align as closely as possible with colors associated with the company, while still being differentiable from each other. Microsoft's XBox line is portrayed in it's iconic green, while Sony's PlayStations are represented with it's later generations dark blue.

As with the scatterplot, this graph also includes a legend. The reasoning is much the same - it allows for quick identification of companies and matches its color. No emojis are present, but squares of color allow the viewer to more quickly make the color comparison, without having to squint at the text color.

LINE GRAPH



This graph is one that depicts the story of the number of copies sold by genre per console. This graph acts as the "cumulative graph" from the first two, which breaks down the sales of games by genre, and then further allows the user to see the distribution of sales by console through the interactivity portion.

The marks of this graph are the lines and the circles. The channels are the colors which differ by genre, and the x and y positions of each of the points, which differ depending on the number of sales per year.

Overall, one problem that we ran into was having the same color scale for both genre and consoles for the interactivity portion of this particular chart where we used both. We felt that there simply were not enough colors to encapsulate all of the different categories for genre and console. While we initially thought of doing similar hues (warm colors, cooler colors) for genre and console, this was also not a good design choice because that implied similarity across genre/console.

Our best design solution that we thought would be clean and appropriate to the user would be to place the legends on opposite sides of our graphs in order to create a clear distinction between the two. Thus, there was no chance that the user could accidentally misinterpret it because they are actively engaging with one side of the screen in order to see a deeper aspect of the story. We felt that this would keep the unity across our graphs while also solving for this potential

area of confusion. This graph uses both legends because it is a combination of the above two graphs.

INTERACTIVITY DESIGN RATIONALE

The interactivity in this project allows the viewer to focus down on a specific genre or console company. It reduces the noise from the other data points and tells the story of the video game market, one genre or company at a time.

The interactivity behind the scatterplot was designed with the intent of showing the specifics for each game that ranked within the top 1000 while also showing how each genre's sales have shifted in popularity over the years. The interactivity portion of the graph allows the user to mouse over a certain point and see details of the point plotted such as the specific game name, its publisher, platform, genre, and the exact number of copies sold. Furthermore, the legend on the left side allows the user to mouse over a specific genre and see the points that are associated with that genre only. This portion of the graph allows the user to see the trends of copies sold for specific genres, so we can see the shift in game trend.

For the bar graph, hovering over a particular company on the legend on the right removes all other bars from the graph. It also shifts the company's bars to the floor to more easily see differences in length. This interaction shows the shift behind how much a console "dominated" the industry each year.

The interactivity in the line graph was designed to show the trends in the copies sold by genre. Furthermore, it allowed the user to see the breakdown of the copies sold by console over the year, allowing the user to observe both trends. By hovering over specific consoles, only the sales for that console are shown. Likewise, hovering over the genre legend allows the user to see the sales of each genre individually. We thought that having these two interactivity portions of the graph provided a unified and cohesive story to the user of how comprehensively, the console and genre of the game impacted its sales.

The unity of having the interactivity for all the graphs be on the same legends adds a sense of unity across our visualizations, and also allows the user to see various trends at the same time across different graphs. We thought this would make our overall visualization more interesting rather than having one interactivity per graph.

CONTRIBUTIONS

Sarah and Kevin worked on designing the graphs while Danny and Valeria worked on getting the Git and csv files set up. We split work depending on which graphs needed to be done, and the expected technical difficulty of each graph. Thus, we had both Valeria and Danny work on the bar graph which had the most complex interactivity, and Kevin and Sarah each individually worked on the scatterplot and line graph respectively. Throughout the entire time, we kept

in communication with one another to update our progress. We came together for a final meeting to compile everything and work on the writeup and finalize our design choices.

DANNY

I was in charge of working on the bar graph with Valeria. Together we split up the work and I worked on the axes and added interactivity. Valeria and I slightly pivoted from the original design laid out by Kevin and Sarah because we thought a bar graph was more conventional and easier to understand. So I constructed the left axis and bottom axis and then added in the interactive legend. The user can hover over the legend and instantly isolate just one console's popularity trend over time. I also added in the description of the bar graph. Overall, I spent around 10 hours on the project and the interactive legend took the majority of that time.

KEVIN

I was in charge of designing the initial graph with Sarah. I was then tasked with implementing the initial scatterplot design, as well as the final integration and polish of the overall interface. I designed and implemented several interactive elements for the scatterplot and line graph, as well as modifications to the interactive elements of all graphs during the final integration process. My work included optimising visualization load times down from seconds to imperceptible, rewiring interactive components to work across graphs, and formatting the webpage for submission. I've spent a little over 10 hours on the project.

SARAH

I was in charge of designing the graphs initially with Kevin. Furthermore, I was also in charge of the line graph visualization with the button animations. The part that took the longest for me was first filtering and reformatting the data to get it to the current form. The data was originally organized by ranking of the game sales, with specifics about the game such as its sales in specific countries and the year it was published. However, to fit the needs of my visualization I reformatted it so that I could extract data from columns with the sales per year for each genre and console. Also, getting the interactivity to work with the buttons took the longest, especially getting the chart to update each time a button was pressed. The time I spent around this project was around 10 hours.

VALERIA

I set up the basic framework of the project, including importing the data. Danny and I also created the bar graph. We discussed design modifications and I processed the relevant data, set up the svg, created the graph, integrated with the axes, and created functions that could alter the graph for interactivity. I also explained our data in our final report and provided more detail on the static design rationale, as well as editing. Approximately 10 hours were spent on this project.