Overview and Motivation:

We chose our dataset out of an interest to explore data relating to the fast-growing and ever-relevant gaming industry. All three of our group members have enjoyed gaming as a hobby from a young age, so it is something that personally interests all of us. This dataset is a great fit for this project, since it provides a massive number (~16,000) data points each representing a released game ranging from 1980 to 2020 and detailing a number of interesting attributes such as platform, genre, release year, publisher, etc. With a large dataset and these useful attributes, there are many interesting visualizations that can be made from this dataset. Many other datasets interested us while we looked around the provided resources for datasets, but the vast majority of them either had incomplete data, a low amount of data points, or an insufficient set of attributes.

Related Work:

There were a few inspirations for our work. A couple of websites periodically publish reports on the top selling videos. For example, A recent article from Digital Trends reported the top selling video games as of March 25, 2021. As far as visualizations discussed in class, the heatmap in particular stuck out to us. While planning our project, we needed a way to visualize two qualitative attributes and one quantitative attribute. The heatmap allowed us to craft this unique visualization.

Questions:

What games are popular in certain regions of the world? Are there games that are popular in certain regions but not others? Are there games that are popular across the board? How did different game systems sell in different regions? These are the main questions that drove our process. However, during the process of making the visualizations, we were still uncertain about what the actual story that our visualizations would be. During our process, we decided we wanted to narrow down the scope a bit to focus in on a single premise: How do the three major publisher's systems perform across each region? Do different genres perform differently on different systems? Does the same genre in a different region perform better or worse? These are some of the questions we wanted to find the answers to.

Data:

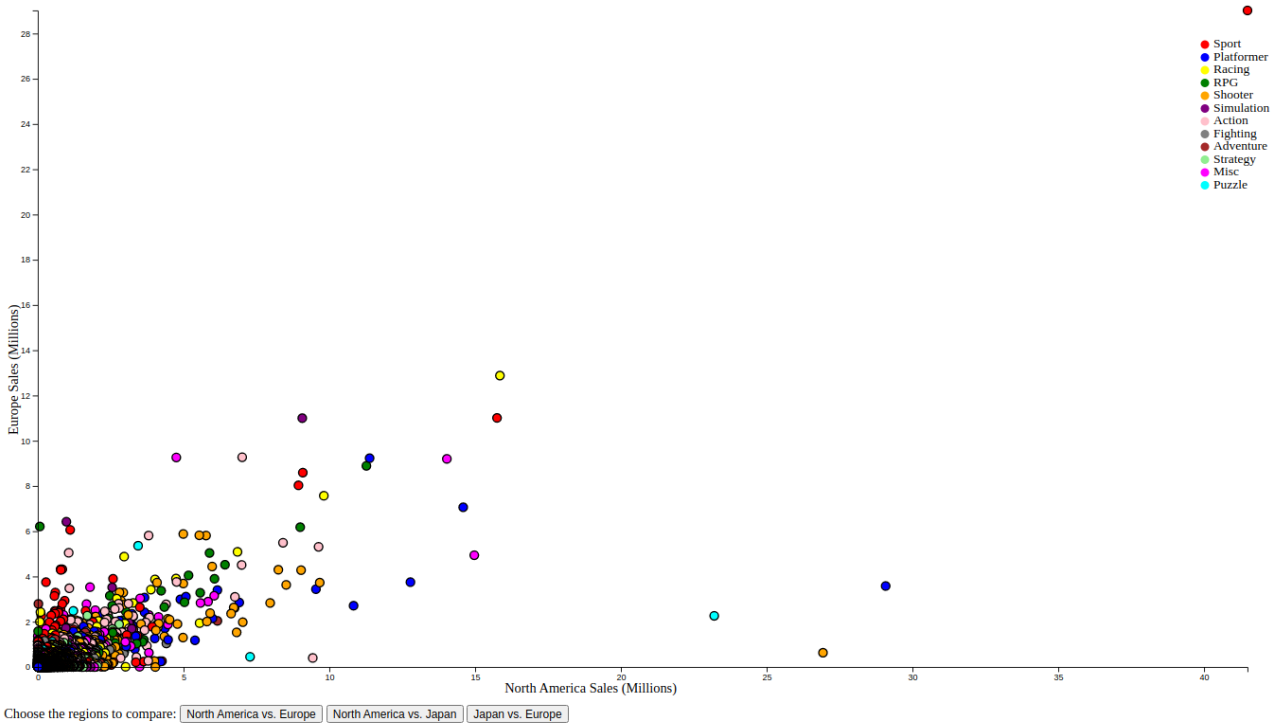
We selected our dataset from here on kaggle.com. It contains a list of games with sales of 100,000 units sold or greater. The dataset was generated from a scrape from vgchartz.com, a video game sales tracking website that provides weekly sales figures of console software and hardware by region. The page includes a csv file containing the actual data we will be creating our visualizations from. For data cleanup, we decided to remove rows that did not include the global sales, as we would have no means to show the game in our charts. For our visualizations, we decided to filter the data based on consoles manufactured by major publishers (Nintendo, Microsoft, Sony). This allowed us to tell a more meaningful story with our visualizations than previously before.

Exploratory Data Analysis:

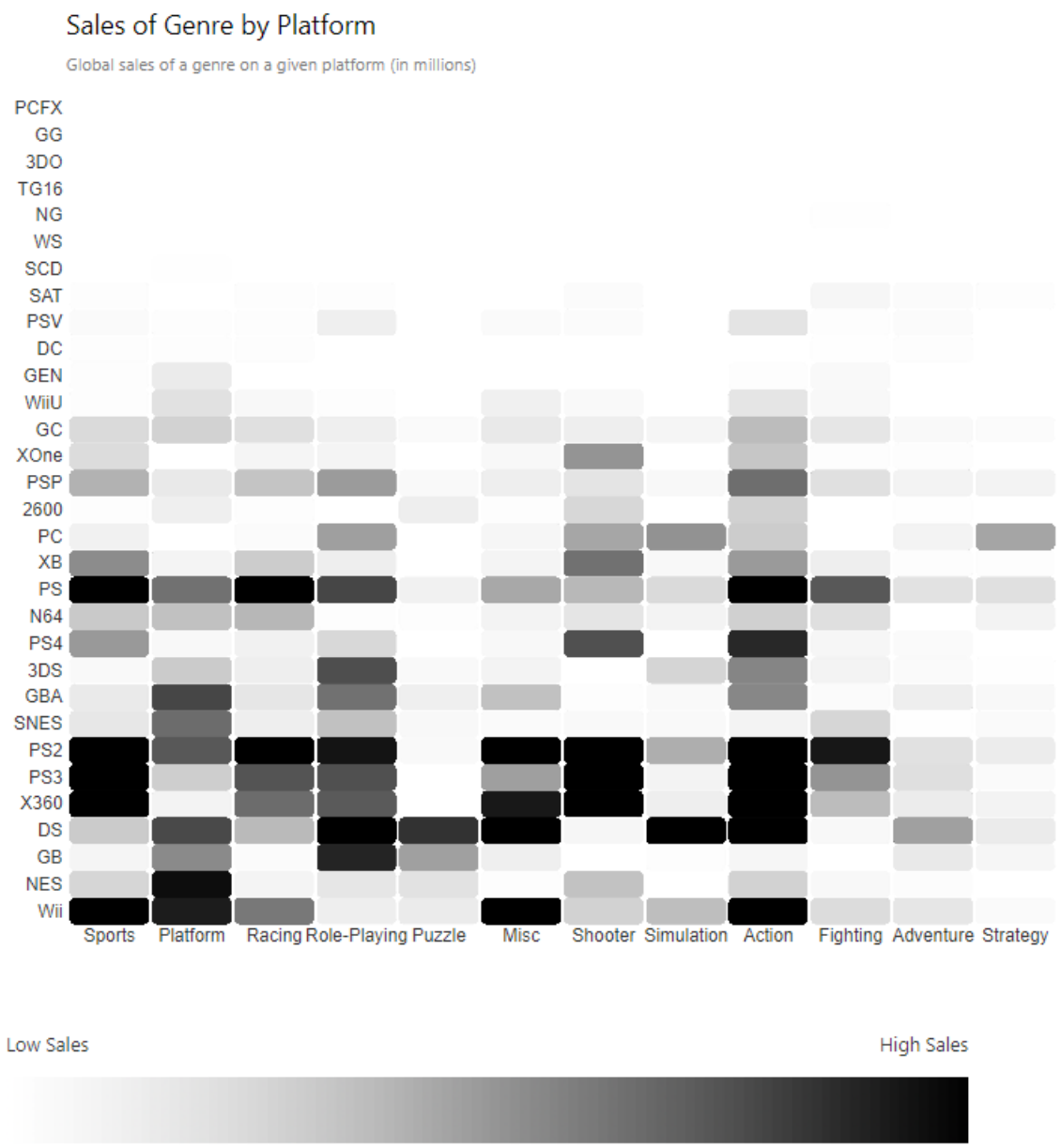
Initially we used a bar chart, line graph and a stacked bar chart to look at our data. This really showed us the impact that different regions had on the sale of video games. Additionally, we noticed that certain publishers were typically selling more games than others. This had a huge impact on our final design. We added select options in order to filter our data by publisher and also optionally filter the data by region.

Design Evolution:

At first, we tried to keep our ideas for visualizations very broad, since we would have to fit so many different publishers and systems together. One of the first designs we thought of was a line chart showing sales over time, but we realized that our dataset could not flesh out this idea too far. Another visualization we tried out was a pie chart showing the market share of different publishers, but it was simply too boring of a design to stick with. As we came up with more and more ideas and implemented them, we realized that the visualizations that worked the best with the dataset were a barchart, a scatterplot, and a heatmap. Only after we realized this and played around with the information it revealed to us did we realize what the story of our project was. For example, this is an early version of the scatterplot visualization:



Here we have every game in our 16000+ sample from our dataset. As you can see, this visualization lacks some clarity, especially with the thousands of clustered lower selling games in the bottom left of the chart. This is when we decided that we needed to find our story so we could provide interesting information people could ask questions about and answer using our visualizations. The idea for the story came after we were analyzing an early version of our heatmap visualization:

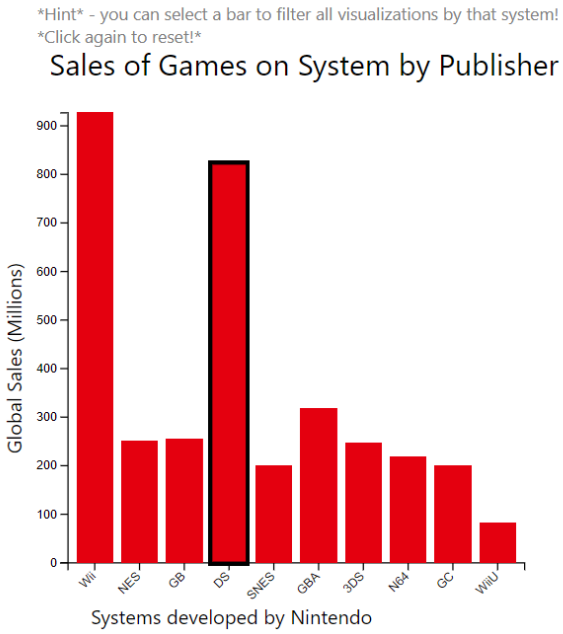


While looking at this, we really wanted to narrow down the game systems we were looking at so you could learn more from the visualization. The story jumped out to us at this moment. From here on out, the goal of our visualization was to investigate how the three major game systems creators (Nintendo, Sony, and Microsoft), performed in different regions, and how certain genres performed on those systems. Having this story fixed a lot of other issues we did not even realize we had. For example, the overall visualization did not have a consistent color

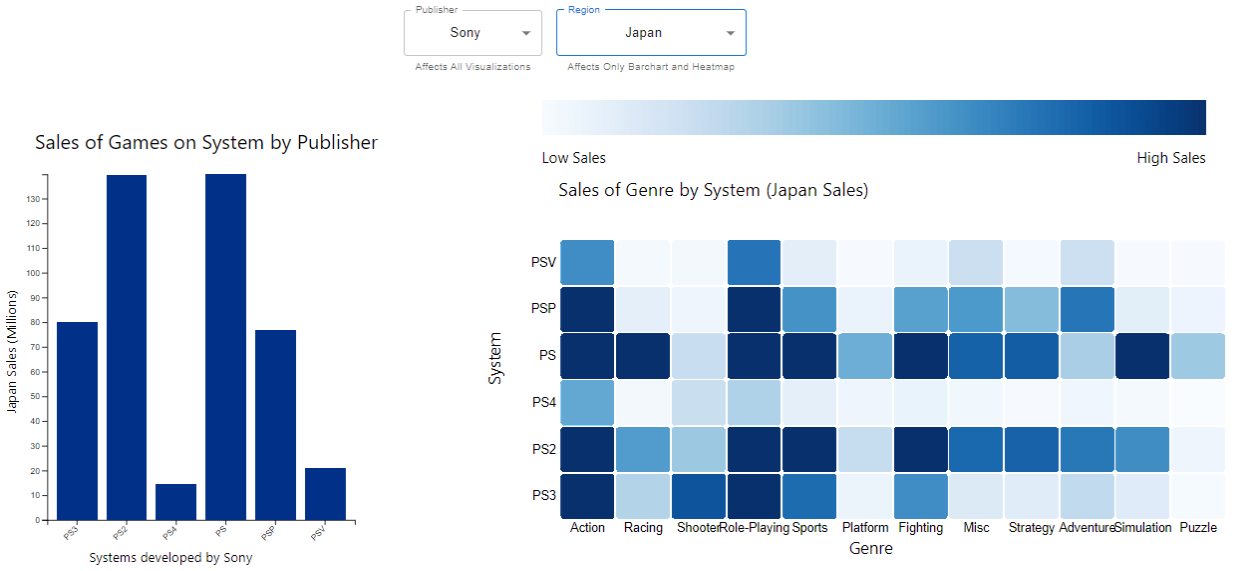
theming. It was a bit all over the place. But by focusing on these major publishers, we could use their own personal branding as a color scheme to differentiate them and the different visualizations. Everything really came together once we had this idea.

Implementation:

For each visualization we have a couple of interactions. For the barchart, the user is able to change the publisher console that the chart visualizes. Additionally the user is able to hover over the chart and view the sales of the system and the top selling game for that system. The user can select a bar and filter all visualizations by the bar chosen.



For the heatmap, you are able to hover over cells in order to get a better view of the data. Also, for both the bar chart and heatmap, you can select which region in order to filter the visualization down with. For the scatterplot, you are able to select the regions in order to compare sales with. Additionally, you are also able to select which publisher in order to display on the view.



Evaluation:

There were several things we learned through these visualizations. One of the things that stood out most immediately was how publishers tended to perform better inside the region that their company is from. For example, Nintendo games sold extremely well in Japan, particularly their role playing games like Pokemon. Microsoft, by contrast, sold abysmally in Japan, while it still had strong numbers in North America and Europe. Another thing we learned is that, while certain genres were popular across systems and regions, certain genres performed better on certain consoles. For example, puzzle games were much more popular on the DS than any other system, especially in Europe and Japan. We think our visualization works well, though of course there are always things we would want to add to it. To further improve our visualization, we would want to have more interaction with the games on the scatterplot. Being able to get further details on those individual games and even having them interact back and forth with the heatmap and barchart instead of only having the barchart and heatmap work on the scatterplot would have been really nice.