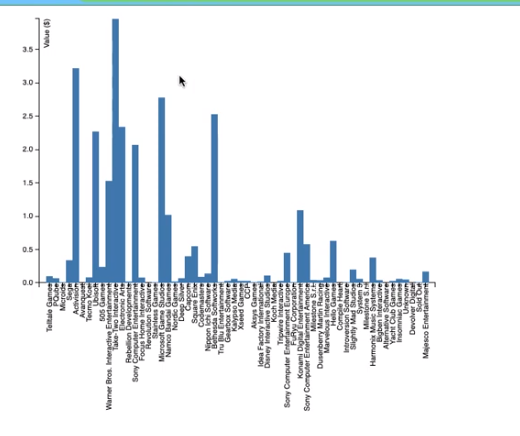
Visualizing Info Final Project

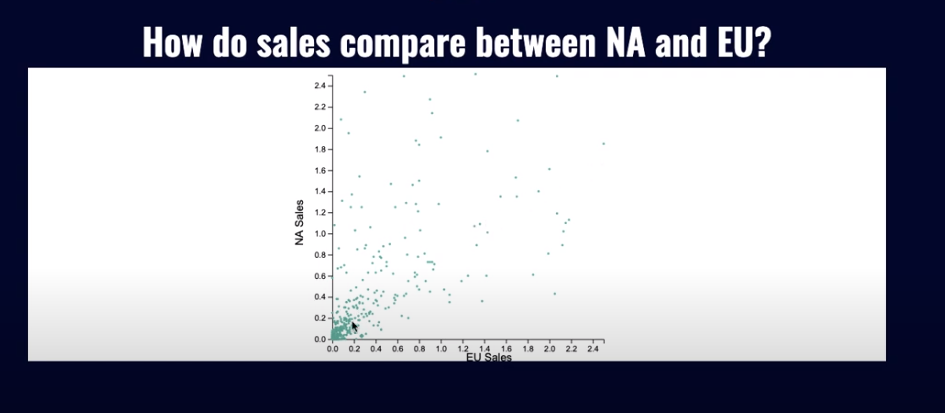
For this project, the dataset that we ended up using is a dataset of video game sales and ratings from 1994 through 2016. It was compiled from a web scrape of VGChartz Video Games Sales and from Metacritic. There were missing observations as Metacritic only covers a subset of the platforms. Also, we found that a game may not have all the observations of the additional variables discussed below. We filtered out the incomplete rows. The total complete rows in this dataset are 6814. We’ve also limited the dataset to the Xbox One and PS4 consoles to make this dataset more manageable. After this filtering, we had a total of 398 rows in the dataset.

For our visualizations we wanted to look at what publisher had the highest total global sales, what publisher had made the most games on Xbox and PS4, and we wanted to look at how a game sold in North America compared to how it sold in Europe. We settled on these questions because we thought that it would be a good way to look at different trends of our dataset to better understand it, and we also wanted to find questions that would allow us to use 3 different types of visualizations.

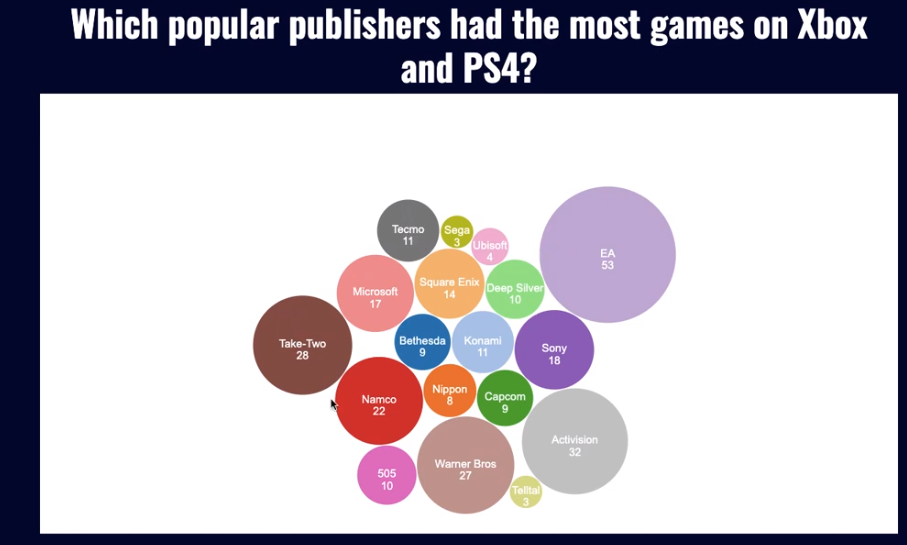
For the first visualization, we asked the question “What publisher had the highest total global sales”. We decided the best visualization to show this would be a bar chart. We thought this would be a good way to easily show the variation of total global sales in millions of units sold for each publisher. This visualization uses bars as the marks and the length of the bars are the channels. The position is also used to show where each bar is alphabetically. We decided to use the same color for all of the bars because we didn’t want the chart to become confusing. From this chart, we could see that Take-Two Interactive was the highest selling publisher by far. It also showed that there were a lot of smaller studios that had very small sales. Overall, this chart did a good job to show us that there were about 20 publishers that had sales that were far higher than the rest.



For the second visualization, we asked the question “How do sales compare between the NA and EU”. At first, we planned on using a bar chart to show the Xbox One sales next to the PS4 sales, but there were far too many video games to be able to do a bar chart. Another reason we changed our minds from the bar chart for this visualization is that we wanted to showcase another type of visualization. We decided to use a scatter plot because it is an easy way to show how well a videogame sold in North America compared to Europe. In this visualization, the marks are points, and the channels are the position of the points. The higher the point is on the chart, the better it sold in North America, and the further right the point is on the chart, the better it sold in Europe. From this chart, it’s easy to see that there are a couple of outliers and there is a positive correlation between the NA and EU sales. From this, we have determined that if a game sold well in North America, that it also sold well in Europe, and vice versa.



For our last visualization, we asked the question “Which popular publishers had the most games on Xbox and PS4”. We also knew that we wanted to use a different visualization than the previous two, so we discussed using either a pie chart or a bubble chart. Either of these visualizations would be able to communicate the information, but we decided that the bubble chart was the better chart to use because it would easily show the ratio difference between total games made by the size of the bubble. The marks for this chart are circles, and the channels are color and size. The color channel has a different color for each of the publishers. The size channel shows how many total videogames the publisher had on the platforms. This is shown by the bigger the circle, the more total games the publisher made, and the smaller the circle, the less games the publisher made. From this visualization, we can tell that EA had more games on XBOX One and PS4 than the other publishers.



These visualizations helped us to understand our dataset better. They showed us that if a game sold well in North America, then it probably sold well in Europe too. We also learned that even though EA had the most games on both platforms, it did not have the highest global sales. The publisher that had the highest global sales is Take-Two Interactive, which had the third highest number of games on Xbox One and PS4. These visualizations helped us to answer our questions, and also gave us other questions to think about for future visualizations. Below you will find our YouTube link that contains a video demonstration of our visualizations.

[YouTube link to visualization demonstration video.](https://www.youtube.com/watch?v=vjWxc5J6AlA)

References

Kirubi, R. (2016, December 30). Video Game Sales with Ratings. Retrieved April 22, 2020, from https://www.kaggle.com/rush4ratio/video-game-sales-with-ratings