I chose my data set because I am really interested in games and found it interesting how the game market has shifted throughout the decade. It was interesting seeing how many games are made by one company or only on one platform because it shows the domination of these companies in this industry. I expected to see lots of PlayStation games and EA games which did happen because many of their games are exclusive to one platform, causing an uptick in that platform game popularity.

My challenge was narrowing down the data because there were so many categorial options, especially for platforms like PlayStation 1, 2, 3, and 4. I overcame these challenges by meshing the platforms with multiple different types into one. Next time I would probably create another variable for the miscellaneous platforms because a big chuck of data was taken out due to constraints. I learned that even if your data set is initially very large, by constraining it, it can become very small very quickly and could cause some bias based on the very specific options.

I focused a lot of the relationship between North American sales and other countries, but I never focused on the other continent variables so if someone else did it, they probably would use all of the continent sale variables. By just focusing on North America, I could have created a little of a bias because I didn’t get the full picture of all the continents’ data.