Background

It is October, 2016, and GameCo's Vice President of Marketing, the Chief Financial Officer, and the Senior Vice President of Sales have asked me to analyze the game sales data they have collected and present insights for 2017 marketing planning.

The GameCo dataset consisted of 16,600 records of units of game software sold to consumers from 1980-2020 (2015 for the purposes of marketing recommendations for the executives). These were categorized by year, region, genre, platform, and publisher. Cleaning of the data involved removing records that didn't list the year or publisher of the game, and averaging the sales of a game across years (imputation) was used to estimate sales in instances when records didn't list sales numbers for a particular year. Based on the cleaning required for this dataset and anticipating future data inquiry, **I recommend collecting sales figures in currency as well as units for greater precision in analysis.**

The executives want to know any insights I can glean from this data, but these questions in particular guide my exploratory data analysis:

- Are certain types of games more popular than others?
- What other publishers will likely be the main competitors in certain markets?
- Have any games decreased or increased in popularity over time?
- How have their sales figures varied between geographic regions over time?

Objective

Perform a descriptive analysis of the GameCo data set to present insights into marketing for 2017, particularly how GameCo's new games might fare in the market in 2017.

Methodology

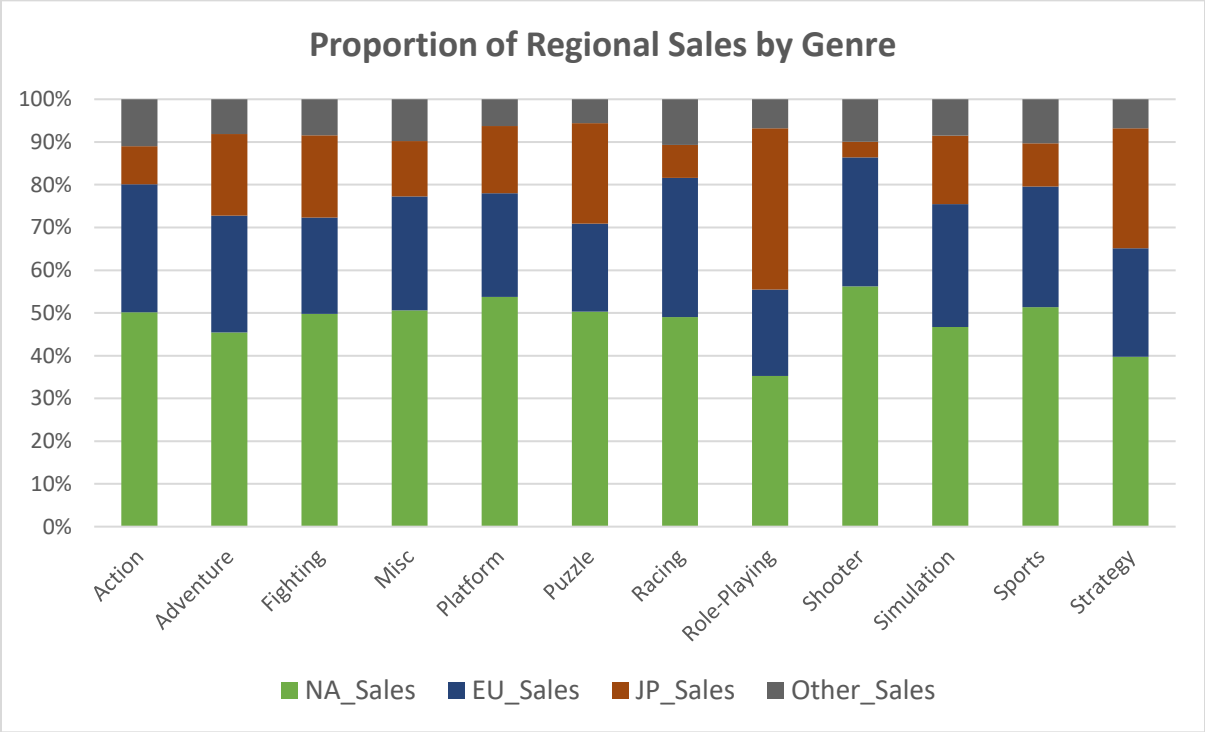
Question: *Are certain types of games more popular than others?*

I used exploratory data analysis to examine popular genres by region, using pivot tables and stacked bar charts. Insights from this analysis:

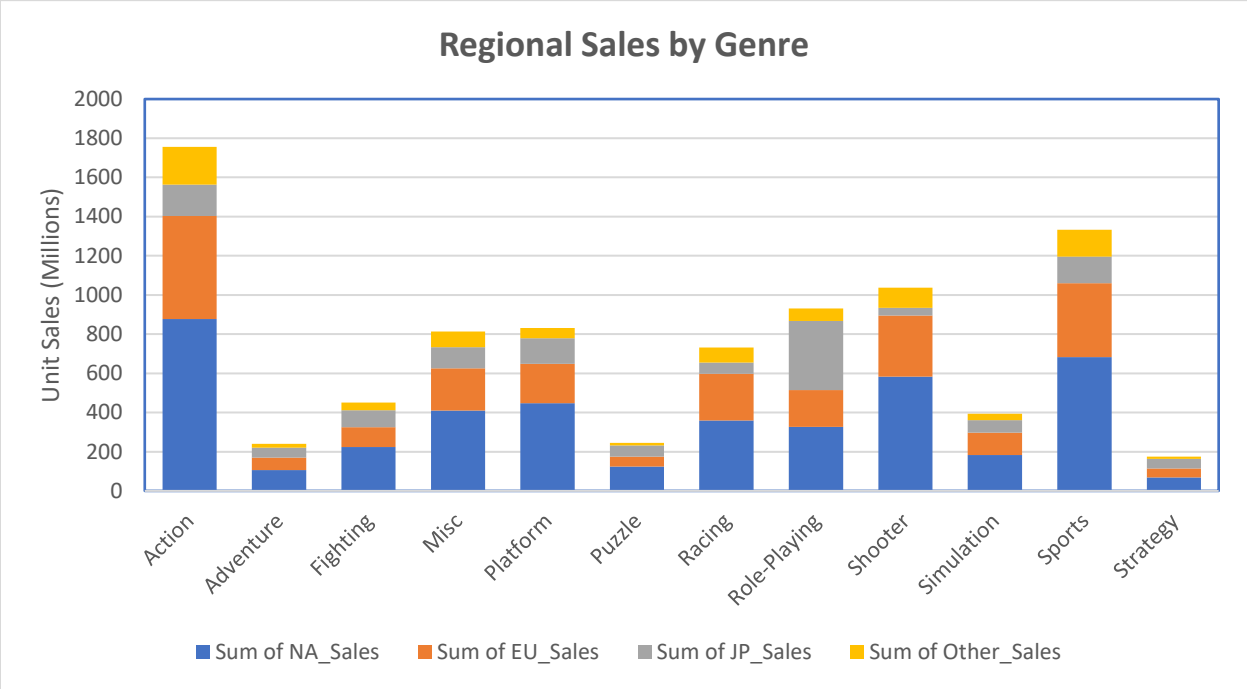
- Genres with the highest sales differ by region when aggregated over all the years for which we have data. Japanese sales are the highest in the genre of Role-Playing, while the other regions are highest in Action.

Genre	EU Sales	JP Sales	NA Sales	Other Sales
Action	525	159.95	877.83	192.79
Adventure	64.1	51.55	105.77	19.22
Fighting	101.32	87.35	223.59	37.88
Misc	215.98	107.76	410.24	79.51
Platform	201.63	130.77	447.05	52.07
Puzzle	50.78	57.31	123.78	13.72
Racing	238.39	56.69	359.42	77.82
Role-Playing	188.06	352.31	327.28	63.22
Shooter	313.27	38.28	582.6	103.09
Simulation	113.38	63.7	183.31	33.65
Sports	376.85	135.37	683.35	136.73
Strategy	45.34	49.46	68.7	12.14
Grand Total	**2435.03**	**1291.3**	**4392.95**	**822.24**

- Role-Playing and Strategy games have the high proportions of Japanese sales at 38% and 28% respectively. The Shooter genre of games have the highest proportion of North American sales at 56%. This insight came from using a pivot table to group sales by genre, using calculated fields to determine proportion of global sales by region, and then visualizing the proportion of regional Sales by genre in a 100% stacked column chart.



- I then did a stacked column chart to compare the actual number of sales, which shows the large differences in sales numbers by genre.



Question: *What other publishers will likely be the main competitors in certain markets?*

I created a pivot table to filter by publisher and percentage of Japanese sales in the global market. This gave insight into the fact that only a few games, publishers, or platforms may account for the high sales in a particular region. This information could indicate publishers that may be competitors in certain markets.

- Japanese sales accounted for 37% of global sales for role-playing games by Ghostlight.

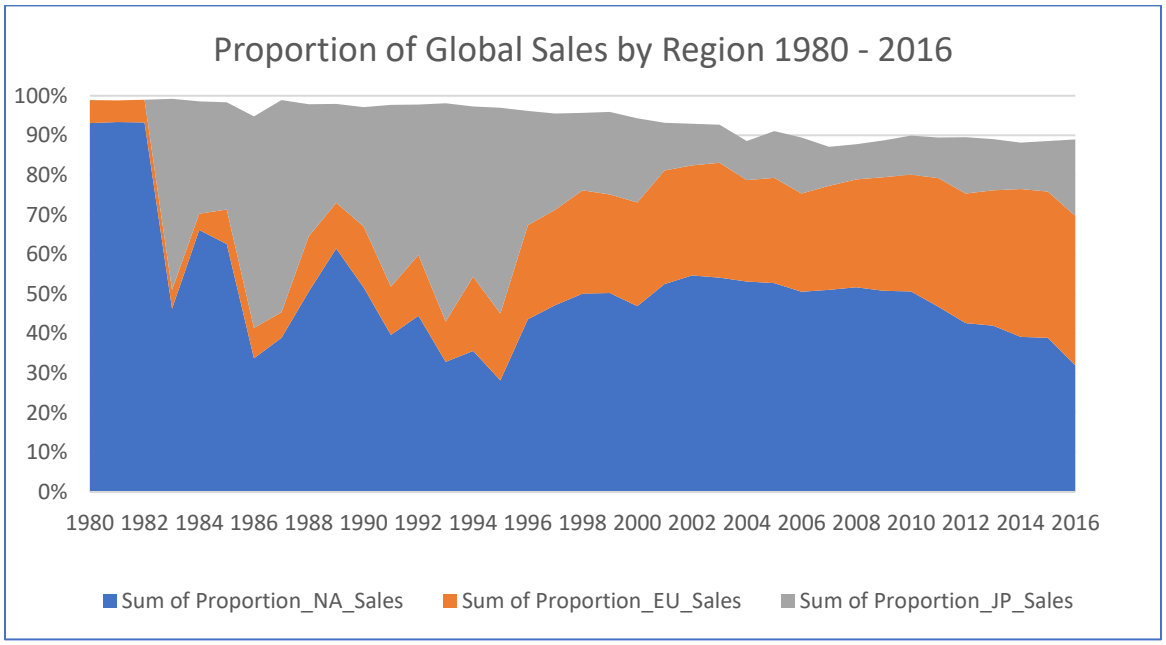
Publisher	Ghostlight	
Genre	(All)	
Sum of JP_Sales	**Sum of Global_Sales**	**Sum of Japanese Percentage of Global Sales**
0.89	2.79	32%

- I also created a pivot table that showed Count of Platform filtered by year and genre. The Action genre has many games available across platforms, Role-Playing has fewer overall. This raises the question about how the quantity of games available affects sales in different regions.

Row Labels	Count of Platform
Action	3316
Adventure	1250
Fighting	848
Misc	1739
Platform	886
Puzzle	582
Racing	1249
Role-Playing	1488
Shooter	1310
Simulation	867
Sports	2346
Strategy	681
Grand Total	**16577**

Question: *How have their sales figures varied between geographic regions over time?*

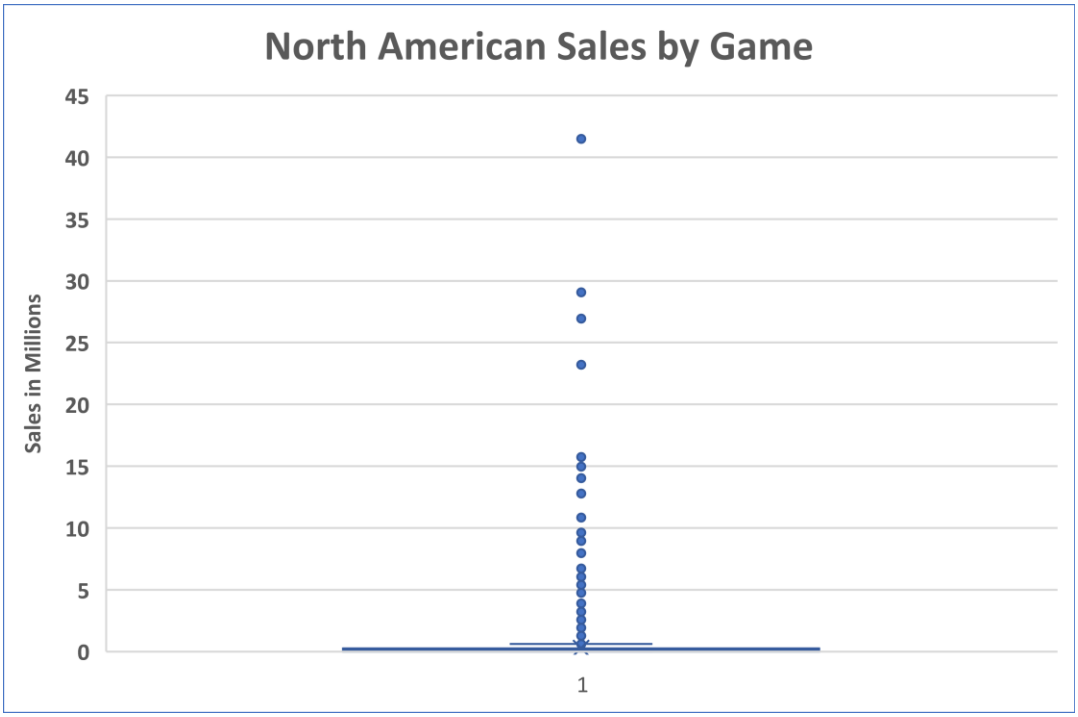
I created an area chart to compare the proportion of global sales by region, which gave insight into how the different regions have changed over time. Seen here, NA sales proportion had been steadily falling for the last 10 years while EU sales proportion was on the rise, and JP sales proportion was mixed over that same time frame.

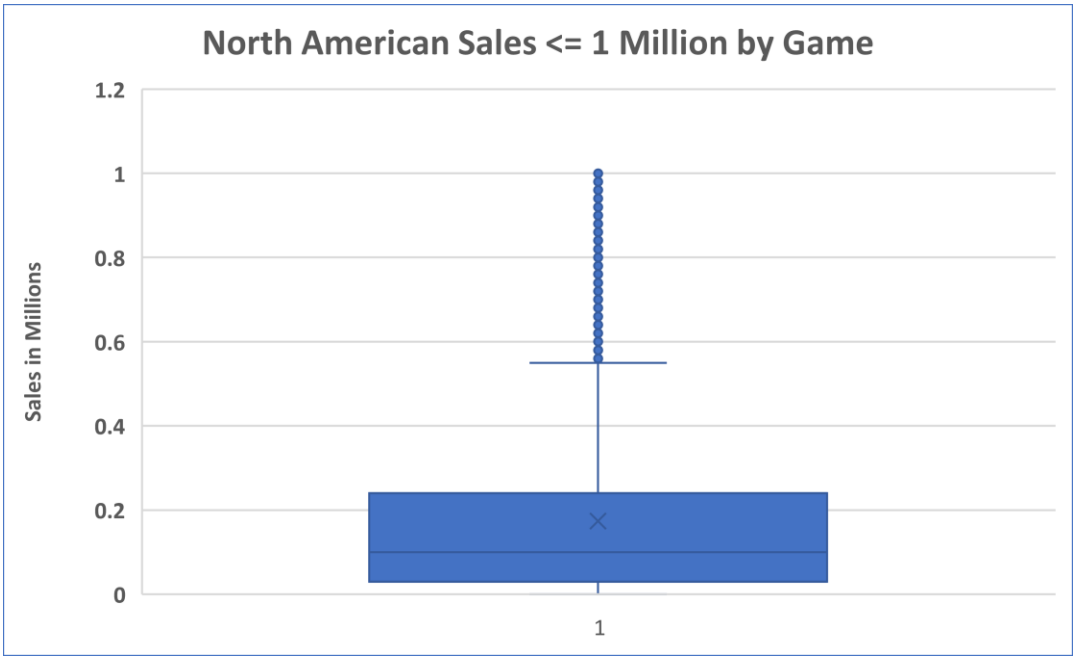


In looking at the sales and proportion figures, however, I recognized that there were very large ranges in the sales numbers. **I calculated and compared mean, median, and IQR for regional and global sales.** This provided insight into the importance of paying attention to what outliers do to the sales numbers. The mean values overrepresent outlier values; indeed, the mode value for all regions is 0.

	NA_Sales	EU_Sales	JP_Sales	Other_Sales	Global_Sales
Mean	0.2646674	0.146706	0.077799	0.049538499	0.5376214
Median	0.08	0.02	0	0.01	0.17
Mode	0	0	0	0	0.02

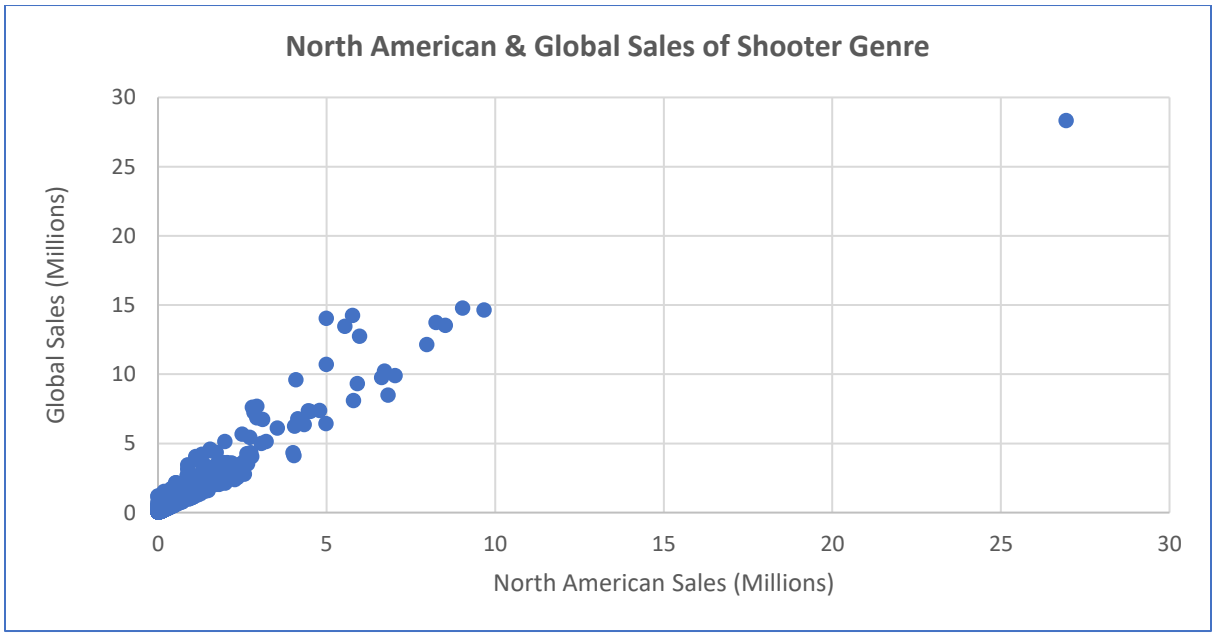
I also created box and whisker plots to visualize North American game sales, using all the sales numbers and then just those under one million. This is additional insight into the effect of outliers on the sales numbers. If I only look at the mean of sales across games, games that represent outlier values of sales will be overrepresented. The vast majority of games had sales much lower than the mean, as seen in the median value visualized here – 78% of games sold less than 1 million. I want to ensure I don't make erroneous conclusions about market and sales patterns without considering the disparity between mean and median for regional and global sales.





Question: *Have any games decreased or increased in popularity over time?*

I created a scatterplot that showed a positive correlation when comparing the relationship between global and NA sales. North American sales are 56% of global sales of shooter games. The North American sales strongly affect the global sales figures.



This correlation, along with my earlier box and whisker charts, provided insight into this final question. If I define “popularity” of a game by its sales, rising and falling sales of particular games and genres may indicate increased or decreased popularity. In the scatterplot, fluctuations in sales of even a few high-selling games would have a large impact on overall sales. This insight, taken with the box and whisker charts from the previous question, provides further indication that outliers are playing a large role in my sales numbers regionally and globally. A deeper dive into these aspects of the data will help me better describe previous trends and make more specific recommendations for marketing in 2017.

Conclusions

My exploratory data analysis revealed the strong influence that high sales of a small proportion of games are having in total sales. I did a deeper analysis to explore the trends we may be missing by looking at total sales rather than accounting for the overrepresented effect these outliers have. From this, I was able to identify the average sales numbers that are more representative of our competitors’ games than the total global sales figures.

Recommendations

My recommendations for the 2017 marketing strategy include using a higher marketing budget for the games from genres that have the highest average sales per game globally as well as high proportions of sales across regions.