

Comparing Profitability for Game Developers: Xbox vs PlayStation vs PC

Thomas Shomer

Introduction

One of my favorite activities to do is playing video games. Ever since the first video games such as Pong and Tetris were created, the gaming industry has continuously expanded and has had a huge impact on today's culture. However, to many consumers, it appears there are only a few very successful competing companies, such as Nintendo and Microsoft, and many failures by independent (Indie) developers. From my experience as a gamer, I would like to see a wide variety of good games offered to consumers instead of only Mario and Call Of Duty games. Therefore, to help Indie-game developers succeed with the games they create, I will investigate for which modern-day platform that Indie developers should create their game.

To determine the success of a game, North American sales will be used. However, this is not the amount of total money earned from given video games. This response variable is measured in millions of units sold. Therefore, I did not need to adjust for the varying prices from these games nor need to adjust for inflation.

For this project, I considered the following variables for predicting North American sales (measured in millions of units sold): user score (scores from 0-10 given from MetaCritic's subscribers), critic score (scores from 0-100 given from MetaCritic staff), ESRB ratings, and genre of the video game.

Data Collection

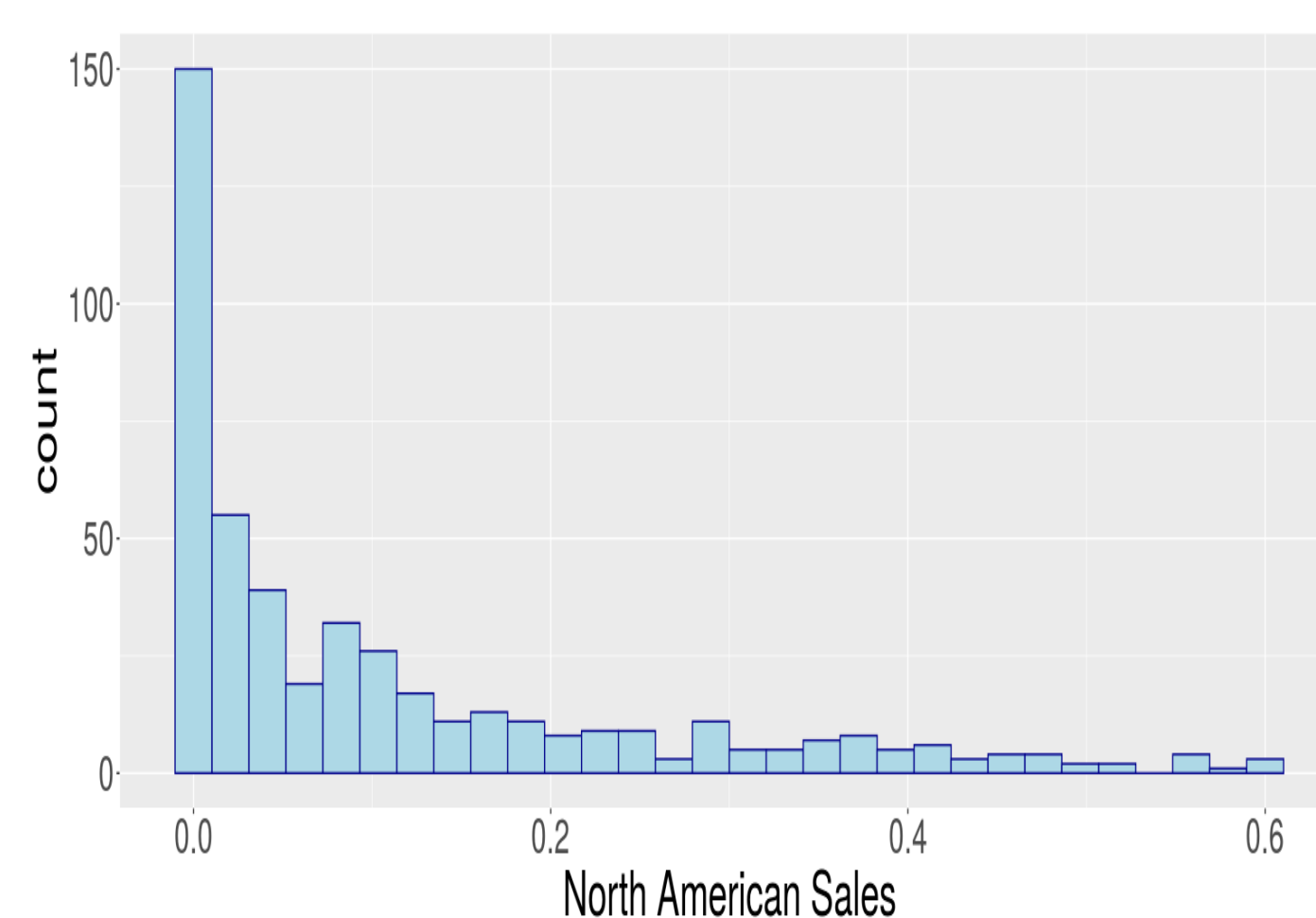
The data that I explicitly found came from one spreadsheet from Kaggle.com. The user scores and critic scores came from MetaCritic, a website which aggregates reviews for many forms of media, one of which is video games. This dataset does not explain where it received the data for sales of all the video games.

This dataset included only games from prominent companies, so I had to adjust this dataset for if I was looking at Indie games. I decided to use the sales of one of the most well-known Indie games as an official cut-off: Undertale. Since this game is an outlier in the Indie-game community, I counted the sales from the year of its release, which I rounded up to 600,000 copies. The assumption that my dataset resembles the distribution of sales for Indie games is much more reasonable.

Another main assumption is that these sales represent present-day sales. In my opinion, this assumption is not valid, but since the dataset covered sales of video games released only up to 2016, it is a key assumption to mention.

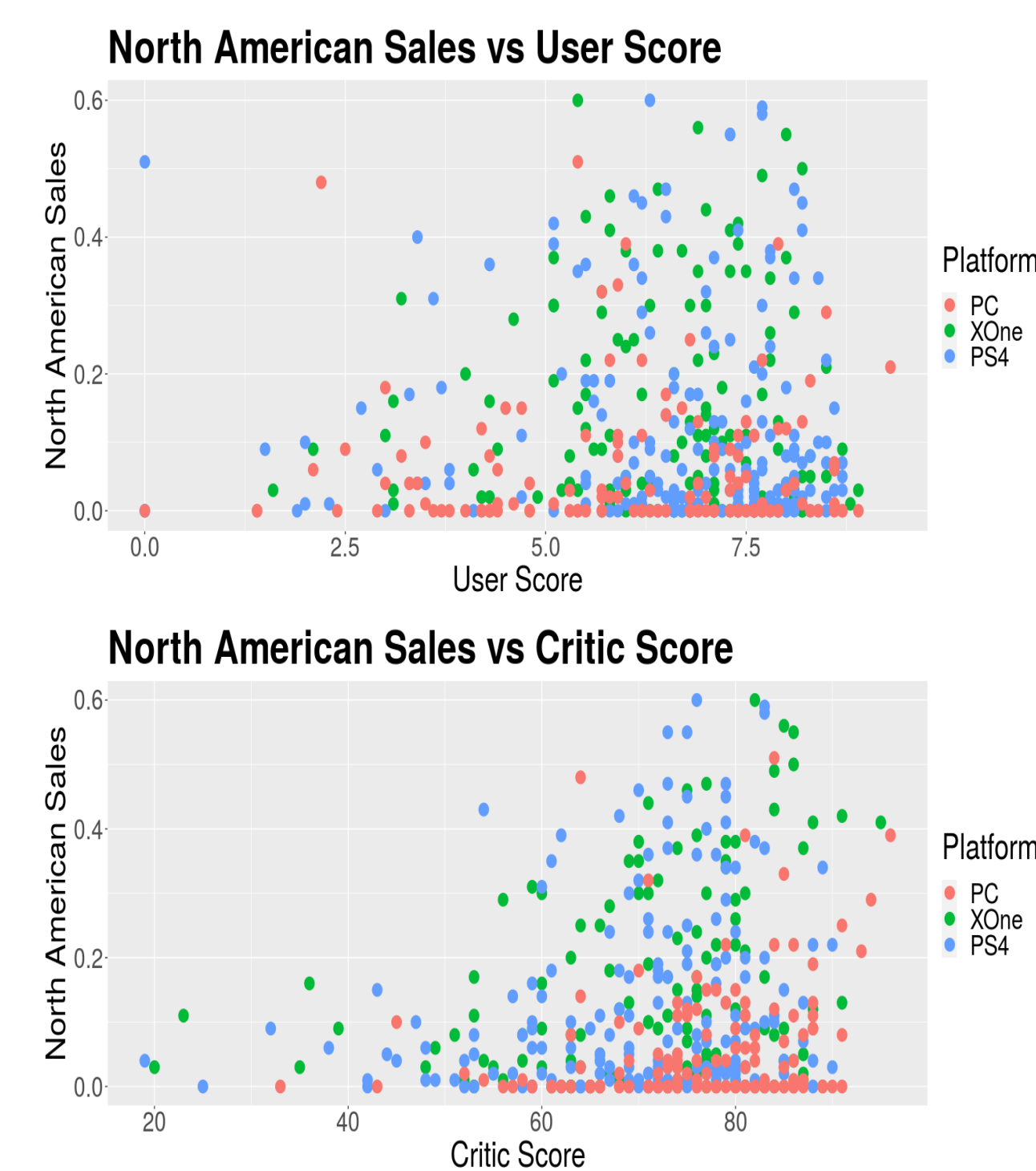
Furthermore, this data was collected near the end of 2016. This biases recent games from 2016 to have less sales than older games. I assume that this bias has an equal effect on individual games if they are of the same year, but with the unequal amount of games between the platforms, it is difficult to assume an equal effect between the platforms. After further adjustments for missing data, my research accounted for 472 total video games from 2013-2016.

Distribution of North American Sales



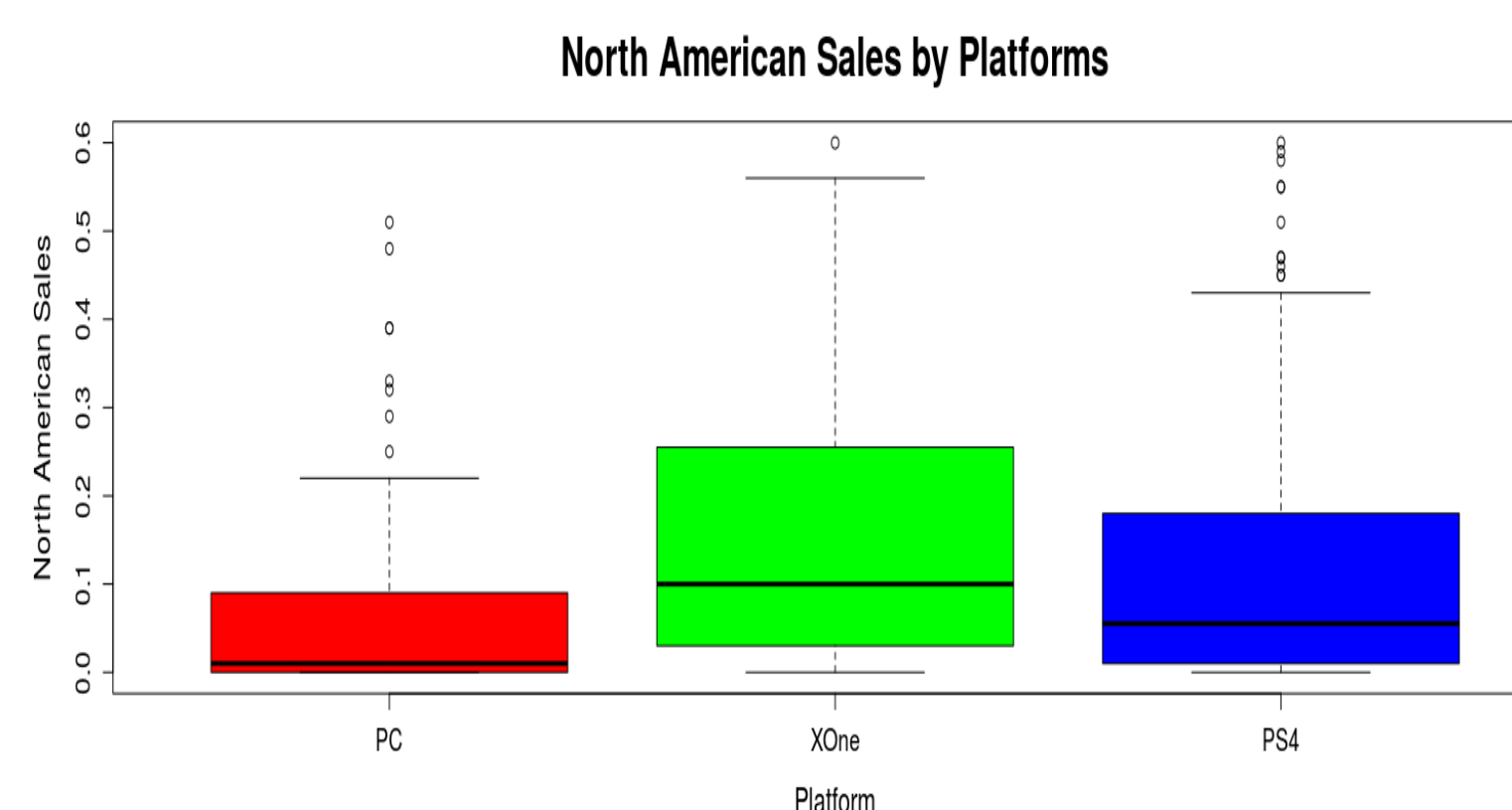
The histogram visualizing North American Sales between 2013 and 2016 appears to not be close to normally-distributed. After outliers and missing values were removed, the data is very positively skewed. The mean of sales is 111,843 and the median of sales is 50,000.

Quantitative Predictor Variables

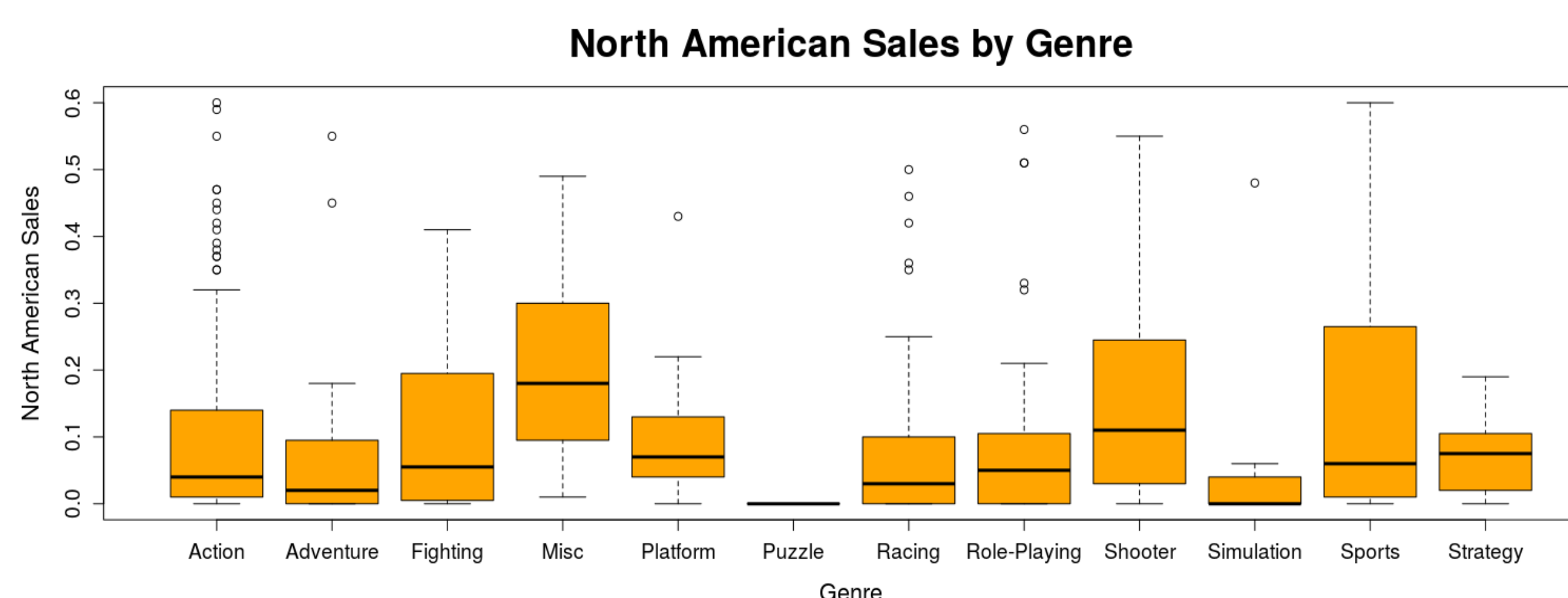


The scatterplots to the left display North American Sales versus user score (top) and critic score (bottom), separated by color to indicate the platform of the video game (red = PC, green = XOne, blue = PS4). Both scatterplots have a very slight positive direction, indicating that sales tend to increase slightly as the user score and critic score increase. The p-value and R² for the user score variable are 0.919 and 0.0022%, respectively. The p-value and R² for the critic score variable are 2.957*10⁻⁵ and 3.9%, respectively. Therefore, the critic score is a stronger predictor of sales than the user score. However, I decided to still include both in my multiple regression analysis.

Categorical Predictor Variables



The side-by-side box plots on the left display the relationship between North American Sales and the three core gaming platforms. All distributions have a positive skew. Xbox One and PS4 sales have a median that appears significantly greater than the median PC sales.



The side-by-side box plots above give insight into the relationship between North American sales and the genre of games being played: Action, Adventure, Fighting, Misc., Platforming, Puzzle, Racing, Role-Playing, Shooter, Simulation, Sports, and Strategy.

There are many different genres provided in my dataset but attempting to account for all of them would be unnecessary and inefficient. It is not immediately apparent by looking at these boxplots, but there are not many data points within every genre that would generate meaningful analysis (ex: Puzzle games). Furthermore, many genres do overlap a lot but only one genre was chosen for each game in my dataset. For example, many Role-playing games could also be considered Strategy or Adventure games.

The other categorical variable I used was the ESRB rating of video games. The 4 main ratings under this variable are everyone (E), everyone 10 years and older (E10+), teen (T), and mature (M). Furthermore, the order of the ratings in terms of highest median North American sales is: E10+, M, T, and E.

Regression Results

I created a multiple linear regression model to predict North American Sales using 9 predictors:

Variable	Coefficient	P-value
PlatformPS4 (1 if PS4; 0 otherwise)	0.0962	<0.0001
PlatformXOne (1 if XOne; 0 otherwise)	0.1265	<0.0001
genreSports (1 if Sports; 0 otherwise)	0.0754	0.0002
genreMisc (1 if Misc; 0 otherwise)	0.0586	0.0505
RatingE10+ (1 if E10+; 0 otherwise)	.0951	<0.0001
RatingT (1 if T; 0 otherwise)	0.0440	0.0152
RatingM (1 if M; 0 otherwise)	0.1133	<0.0001
Critic Score	0.0041	<0.0001
User Score	-0.0174	0.0003

The genre variables were chosen with best subsets regression. The ESRB rating of E was chosen to be the default category because it is the default ESRB rating for a given game unless other age-restricted content is present. The PC platform was chosen to be the default category since it had the lowest median sales.

Interpretation of Results:

- Compared to a PC game, holding other variables constant, a PS4 game is predicted to sell 96,200 more copies in North America.
- Compared to a PC game, holding other variables constant, an Xbox One game is predicted to sell 126,500 more copies in North America.
- Compared to games of other genres, holding other variables constant, a sports game is predicted to sell 75,400 more copies in North America.
- Compared to games of other genres, holding other variables constant, a miscellaneous game is predicted to sell 58,600 more copies in North America. However, this variable is arguably not significant.
- Compared to games with an E rating, holding other variables constant, an E10+-rated game is predicted to sell 95,100 more copies in North America.
- Compared to games with an E rating, holding other variables constant, a T-rated game is predicted to sell 44,000 more copies in North America.
- Compared to games with an E rating, holding other variables constant, an M-rated game is predicted to sell 113,300 more copies in North America.
- For an increase of 1 in critic score, holding other variables constant, a game is predicted to sell 4,100 more copies in North America
- For an increase of 1 in user score, holding other variables constant, a game is predicted to sell 17,400 less copies in North America

With an R² value of 0.2695, this model accounts for 26.95% of the variation in North American sales from 2013 to 2016.

Conclusions and Future Work

Through my multiple-regression analysis, I determined that Xbox One and PS4 are both great platforms for an Indie developer, but the PC is significantly worse compared to both the Xbox One and PS4.

In the future, I hope to have much more recent data that is as complete as the one I used for this project. Due to my dataset being almost 4 years old, it does not portray recent developments for PC gaming. I would think that, if given more recent data, PC would be significantly better than both Xbox One and PS4 instead of worse. Another assumption I hope to adjust for is that there is an equal spread of games for all years between platforms. For example, if PS4 had more 2016 games than Xbox One, and Xbox One had more 2013 games than PS4, then Xbox One sales could be unfairly higher than PS4.

Furthermore, I hope to analyze more sales instead of just in North America. European sales, Japan sales, and global sales are included in my dataset, but the scope of this current project could not include analyzing those sales along with North American sales.