

# Correlation Between Video Games & Critic, User Scores



Ryan Lee

# Data Driven Blog

<https://www.kaggle.com/datasets/rush4ratio/video-game-sales-with-ratings>

# Overview

I am senior majoring in computer science. I was always interested in video games and I am currently seeking for a job in video game field. After looking at this assignment, I was wondering if the critic&user score actually affects to the global sale of the video games. Is specific genre's game sold more than others? We can find it through z-test.

## Dataset

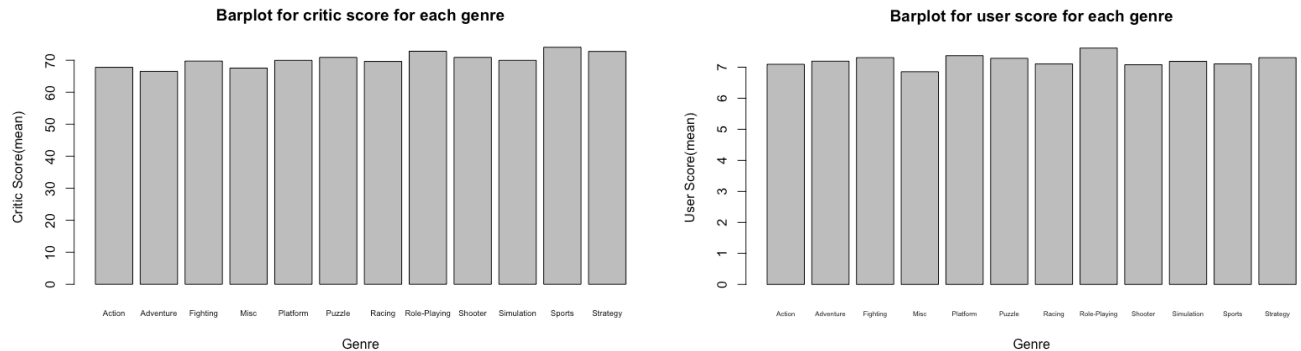
Fortunately, I could find the dataset from Kaggle to analyze about it. Game sales records are from [VGChartz.com](http://VGChartz.com). They share sales record of each game and write column about comparison for game platforms. The score is from [metacritic.com](http://metacritic.com). They are popular for rating movies and game. This dataset only includes games with sales greater than 100,000 copies. The purpose of this dataset is to use these data for machine learning technique or continued data visualizations. We will look for some data trends with this dataset. I think this dataset is trustful because metacritic is one of the biggest rating website for the movies & games and only verified critics can leave their ratings.

## Pre-processing

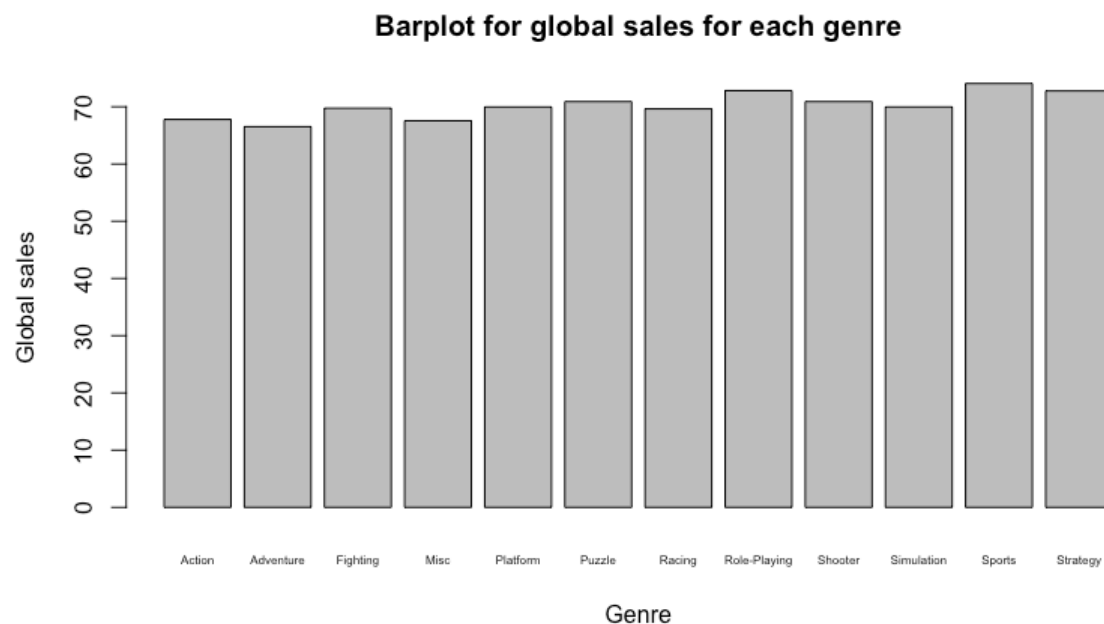
```
1 video <- read.csv('Video_Games_Sales_as_at_22_Dec_2016.csv')
2 video <- na.omit(video)
3 video$User_Score = as.numeric(video$User_Score)
```

First, I dropped the rows that have null values in the dataset. And User\_Score variable was saved as “character”, so I changed it to numeric value. Now, we are good to use dataset for plots and tests.

# Plots for Dataset

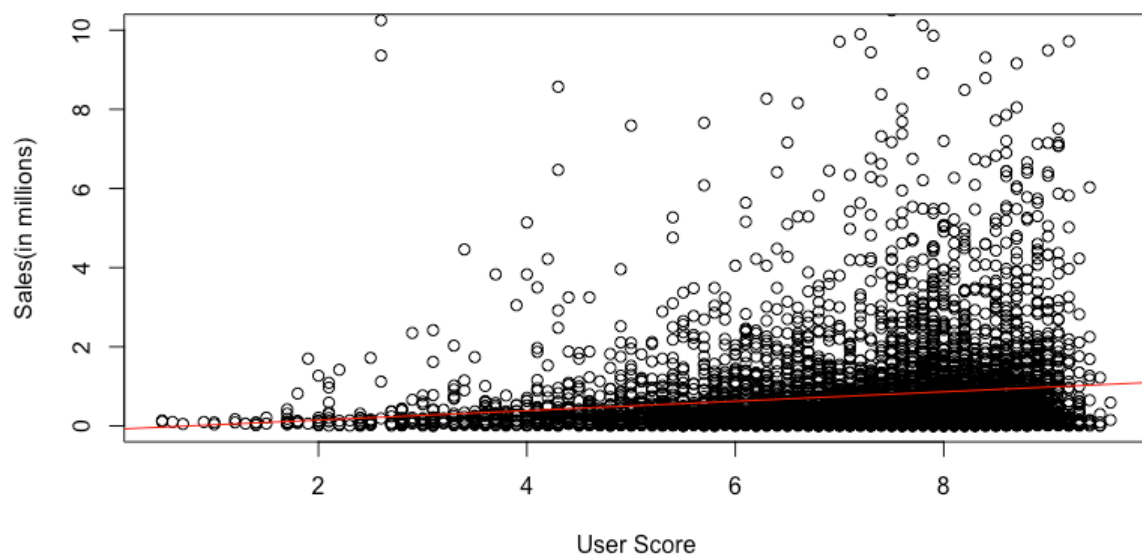
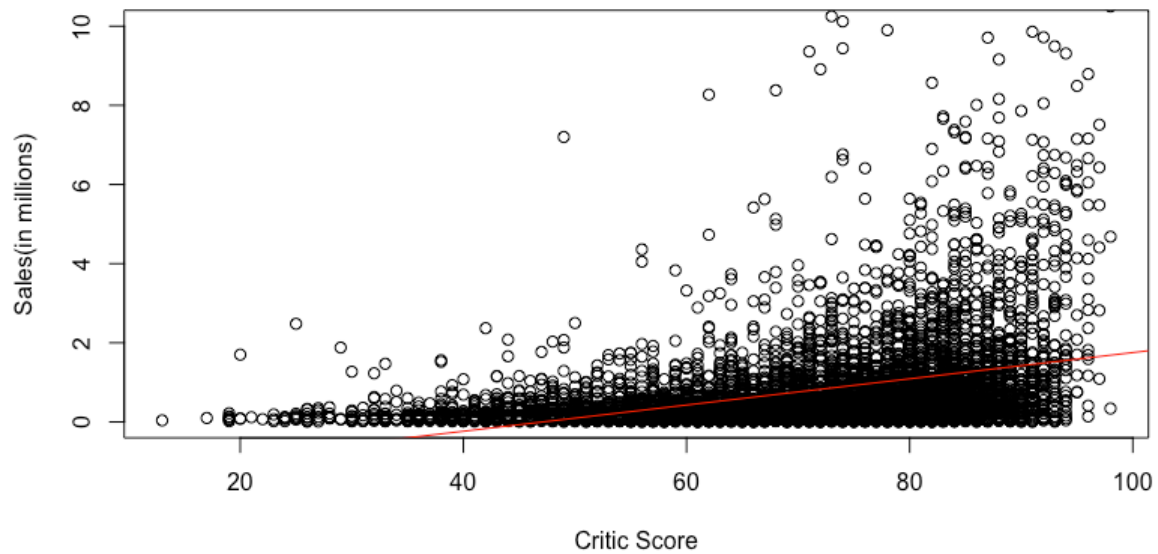


From this bar plot, we can see Role-Playing and Sports genres' critic score is generally higher than the others. And Role-Playing genre's user score is generally higher than others. The trend for graph is kinda similar between users and critics, but Sports genre were stronger in critics side while Adventure genre is stronger in users side.



We can see Role-Playing and Sports games were generally sold more than other genres with this bar plot. So, we can assume the score and sales are somewhat proportional, but this is not enough. So, we go further.

Here is another evidence that score and sales are proportionally related. I made a scatter plot with linear regression line.



The red line shows the trend for the relationship between score and sales. The line is going upward as the line goes right, so we can say it is proportionally related. For last, we will make the hypothesis and run a Z test.

```

video <- transform(video, critic_category = cut(video$Critic_Score, breaks = c(0,50,75,100), labels = c('Low', 'Medium', 'High')))
ZTest::z_test_from_data(video, 'critic_category', 'Global_Sales', 'Medium', 'High')
video <- transform(video, user_category = cut(video$User_Score, breaks = c(0,5,7.5,10), labels = c('Low', 'Medium', 'High')))
ZTest::z_test_from_data(video, 'user_category', 'Global_Sales', 'Medium', 'High')

```

With this code, I classified the score. For critic score, 0~50 is low, 50~75 is medium, and 75~100 is high. For user score, 0~5 is low, 5~7.5 is medium, and 7.5~10 is high.

Null hypothesis is “Games are sold the same regardless of score.”

Alternative hypothesis is “Better scored games will be sold better.”

We conducted a Z-test between Medium score and High score and result are like these.

```

> ZTest::z_test_from_data(video, 'user_category', 'Global_Sales', 'Medium', 'High')
[1] "6.93239594060239 is the z-value"
[1] "2.06890060638898e-12 is the p-value"
[1] 2.068901e-12

```

```

> ZTest::z_test_from_data(video, 'critic_category', 'Global_Sales', 'Medium', 'High')
[1] "15.2453488031421 is the z-value"
[1] "0 is the p-value"
[1] 0

```

## Conclusion

As we can see with the z-test result. Our p-value is less than 0.05, which we can reject the null hypothesis. I showed three evidences to analyze the correlation between score and sales, and we could know better scored games are sold more. So, it would be important to make good games if the company want to sell many games and get some money from it.