A3: Exploratory Data Analysis with Tableau

CSE412

Vicky Yeh, shuyiyeh@uw.edu, AC

Dataset Overview

Introduction
This dataset contains video game sales data as of 2019. There are a lot of information, comprising details about video games, their sales, platforms, genres, critic scores, and more. Additionally, the dataset provides sales information for games across various regions, including North America, Europe, and Japan.

How/Why it was selected
I'm really interested in playing video games across different platforms. Therefore, I'm curious about highest-selling games and their scores, as this could guide me in deciding which games to purchase. This dataset meets all the criteria and provides the comprehensive information I am looking for.

Link
https://www.kaggle.com/datasets/ashaheedq/video-games-sales-2019?select=vgsales-12-4-2019-short.csv

What variables does the dataset contain?
Rank, Name, Platform, Genre, Rating, Publisher, Developer, Critic Score, User Score, Total Shipped, Global Sales, NA Sales, PAL Sales, JP Sales ,Other Sales, Year

How is the data distributed?
The distribution of the user score and critic score are similar, showing a left skewness. There are few outliers which represent games with significantly higher sales than others.

Are there any notable data quality issues?
There is a substantial number of missing data for user scores and ratings. Also, there is one Null data in either the variables 'Total Shipped' or 'Sales' within a game, indicating that these two variables should be integrated into a single variable.

Are there any surprising relationships among the variables?
The trend of NA sales within different gerne is really similar with Global sales, while JP sales shows a total different relationship.

Question List

Question 1
What are the 10 highest sales games in US?

Question 2
What are the relationship between NA sales and global sales?

Question 3
What are the Highest Sales' Developers and their best games?

Question 4
Do specific game genres have different sales proportions in three regions?

Question 5
What is the platform with the highest sales?

Question 6
Is there any difference between user score and critic score within each ESRB rating?

Visualization 1

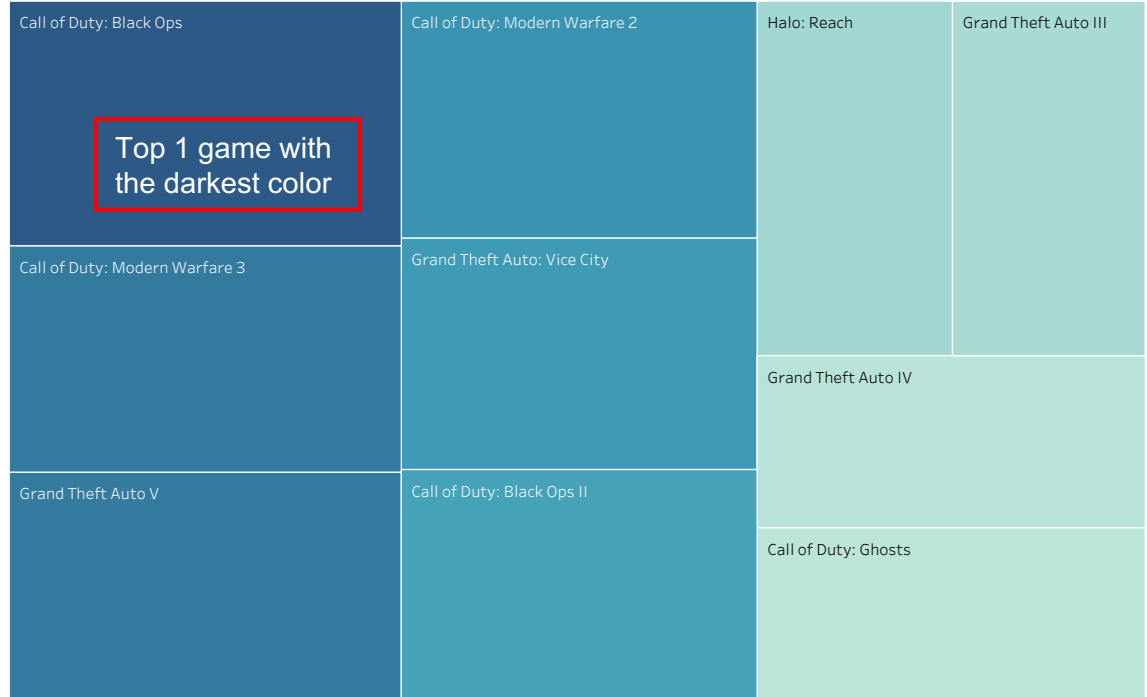## Question 1
What are the 10 highest sales games in US?

Brief description
This visualization uses treemap to show 10 highest sales game in North America (NA). The size of the regions and color both represent the sales of these 10 games.

Key Takeaway
Without actual sales number, we can still understand the ranking of game by observing the color and size.

10 Highest Sale Games in North America

Name. Color shows sum of NA Sales. Size shows sum of NA Sales. The marks are labeled by Name. The data is filtered on NA Sales, which ranges from 6.75 to 9.76.

NA Sales
6.760   9.760

Visualization 2

Question 2
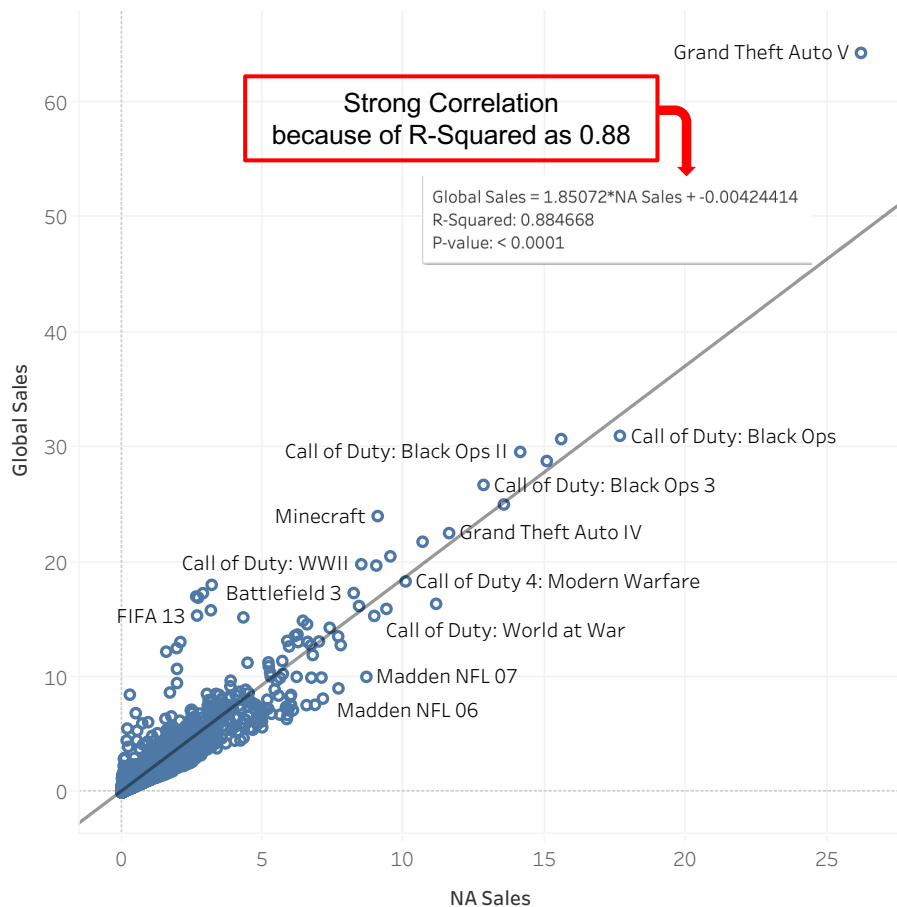What are the relationship between North America sales and global sales?

Brief description
This visualization uses scatter plot to show the correlation between global sales and North America sales. The blue circle represent each game, and the gray line is the trendline of two variables.

Key Takeaway
We can find that Global Sales and NA Sales show a high level of correlation, with an R-Squared value of 0.88. It might shows that games which sell well in the United States tend to be popular globally as well.

Relationship between NA Sales and Global Sales

Strong Correlation because of R-Squared as 0.88

Global Sales = 1.85072*NA Sales + -0.00424414
R-Squared: 0.884668
P-value: < 0.0001

Sum of NA Sales vs. sum of Global Sales. The marks are labeled by Name. The view is filtered on sum of NA Sales, which ranges from 0.00 to 26.19.

Visualization 3

Question 3
What are the Highest Sales' Developers and their best games?

Brief description
This visualization uses horizontal bar chart to show three developer with highest sales and their 5 games with highest critic score. The color is to differentiate the developer more clearly.
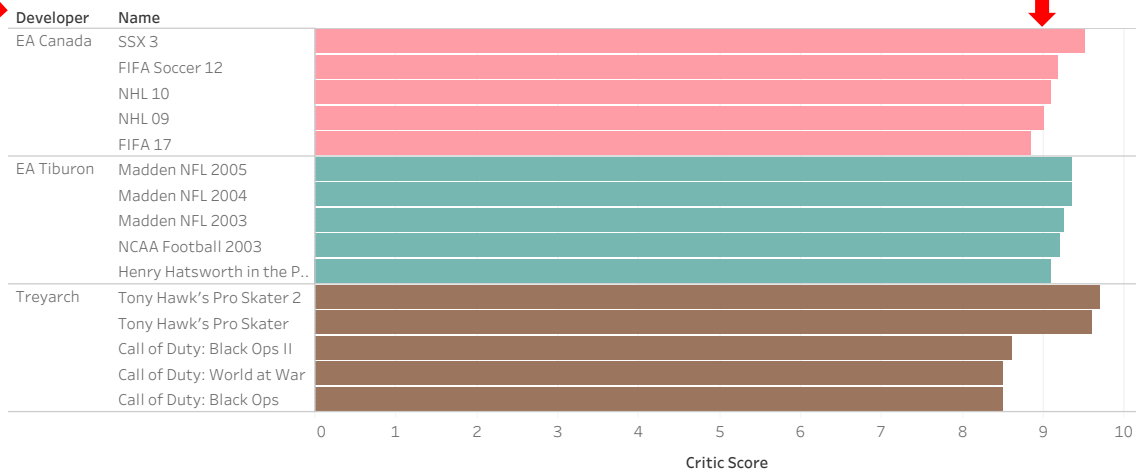
Key Takeaway
- From this figure, we can easily find out the most famous developers and the most popular games from them.
- In stead of focusing on games only, information of developer is also valuable for customers.

Extract the developer with highest sales

Rank the game by critic score

Three Highest Sales' Developer and Their Top 5 Games with Highest Critic Score

Developer	Name
EA Canada	SSX 3
	FIFA Soccer 12
	NHL 10
	NHL 09
	FIFA 17
EA Tiburon	Madden NFL 2005
	Madden NFL 2004
	Madden NFL 2003
	NCAA Football 2003
	Henry Hatsworth in the P..
Treyarch	Tony Hawk's Pro Skater 2
	Tony Hawk's Pro Skater
	Call of Duty: Black Ops II
	Call of Duty: World at War
	Call of Duty: Black Ops

Critic Score

Average of Critic Score for each Name broken down by Developer. Color shows details about Developer. The data is filtered on Critic Score and Index. The Critic Score filter ranges from 7.36 to 10. The Index filter keeps 1, 2, 3, 4 and 5. The view is filtered on Developer, which keeps EA Canada, EA Tiburon and Treyarch.

Developer
- EA Canada
- EA Tiburon
- Treyarch

Use color to differentiate the developer

Visualization 4

Question 3
What are the Highest Sales' Developers and their best games?

Brief description
This visualization uses circle chart to show the same question as previous slide. The position of circle represents the critic score, with the corresponding value indicated below. However, in this chart, we don't use different color to mark anything.

Key Takeaway
- With actual value labeled on the graph, it is more clear for reader.
- It provides a different way to convey the information.

Three Highest Sales' Developer and Their Top 5 Games with Highest Critic Score
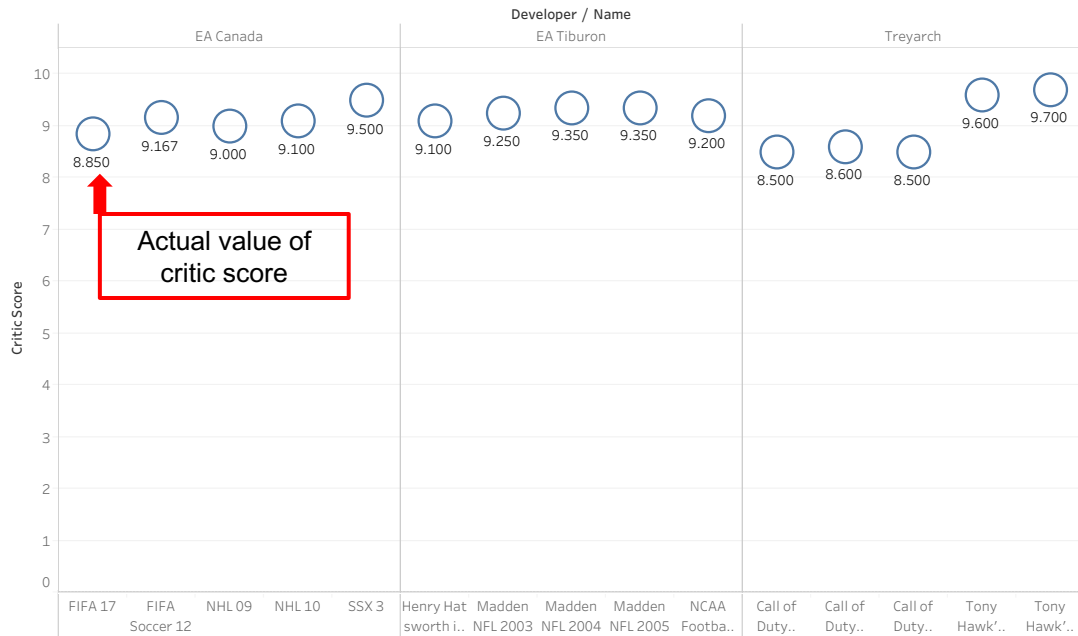
Average of Critic Score for each Name broken down by Developer. The marks are labeled by average of Critic Score. The data is filtered on Critic Score and Index. The Critic Score filter ranges from 7.36 to 10. The Index filter keeps 1, 2, 3, 4 and 5. The view is filtered on Developer, which keeps EA Canada, EA Tiburon and Treyarch.

Visualization 5

Question 4
Do specific game genres have different sales proportions in three regions?
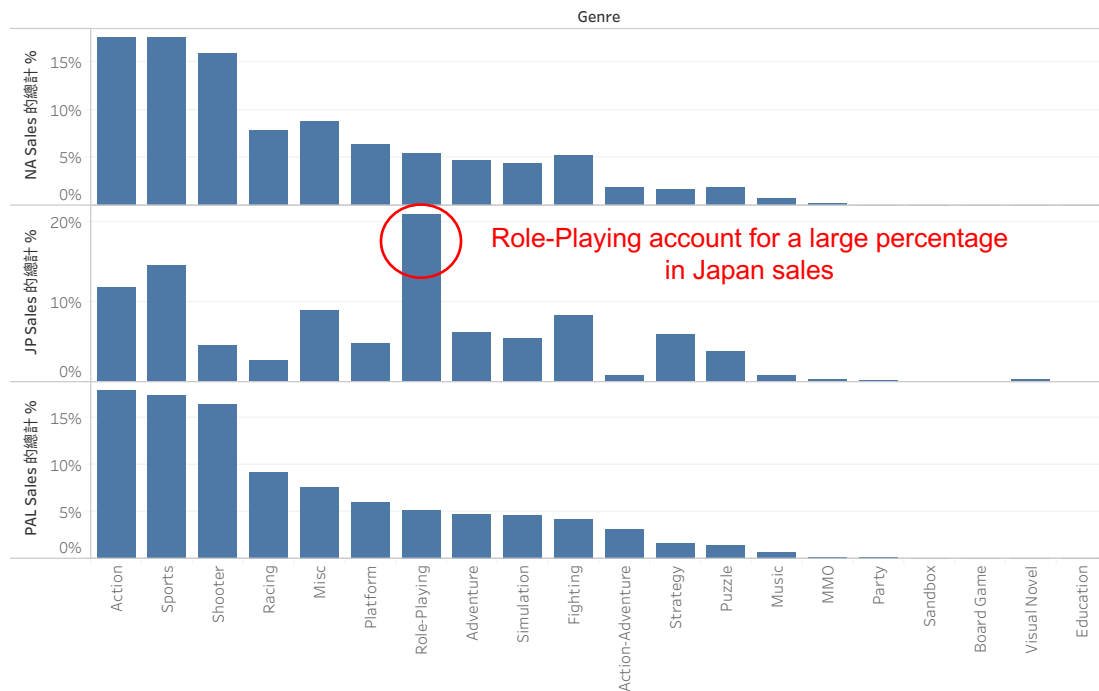
Brief description
This visualization uses 3 bar charts to illustrate the sales from three different regions, which are North America (NA), Japan (JP), and Europe (PAL) respectively. And the sales are further categorized by genre.

Key Takeaway
- We can find that sales of different genres in North America almost has the same distribution as Europe.
- Role-Playing shows much higher percentage in Japan than in other regions.

Comparison of Sales from Different Genres in 3 Regions (NA, JP, and PAL)

Role-Playing account for a large percentage in Japan sales

% of Total NA Sales, % of Total JP Sales and % of Total PAL Sales for each Genre.
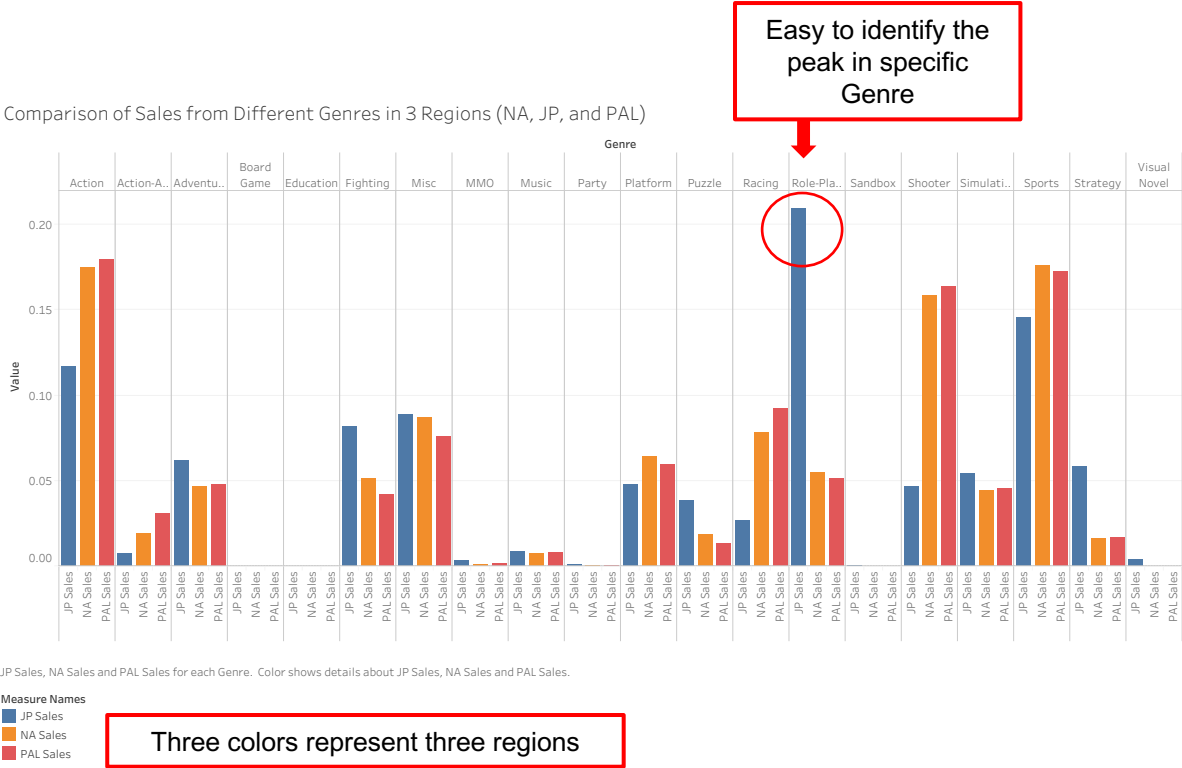
Visualization 6

Question 4
Do specific game genres have different sales proportions in three regions?

Brief description
This chart also utilizes bar charts to answer Q4, but distinguishes different regions using colors. Additionally, it groups the same genres together for comparison.

Key Takeaway
From this visualization, comparing sales across different regions within each genre becomes easier. We can identify significantly high or low values within specific genres more effectively.

Comparison of Sales from Different Genres in 3 Regions (NA, JP, and PAL)

Easy to identify the peak in specific Genre

JP Sales, NA Sales and PAL Sales for each Genre. Color shows details about JP Sales, NA Sales and PAL Sales.

Measure Names
JP Sales
NA Sales
PAL Sales

Three colors represent three regions

Visualization 7

Question 5
What is the platform with the highest sales?

Brief description
This visualization uses packed bubble chart to represent global sales of games on different platform from 2010 to 2020. The size of each bubble corresponds to the number of sales.

Key Takeaway
By observing the size of circle, we can find that game sales on PS3 is the highest, with PS4 and X360 coming in second and third place.

Total Global Sales on Different Platform from Games Released Between 2010 and 2020

The bigger size, the higher sales

X360 — 3rd
PS3 — 1st
PC
WiiU
3DS
DS
Wii
NS
PSV
PSP
PS4 — 2nd
XOne
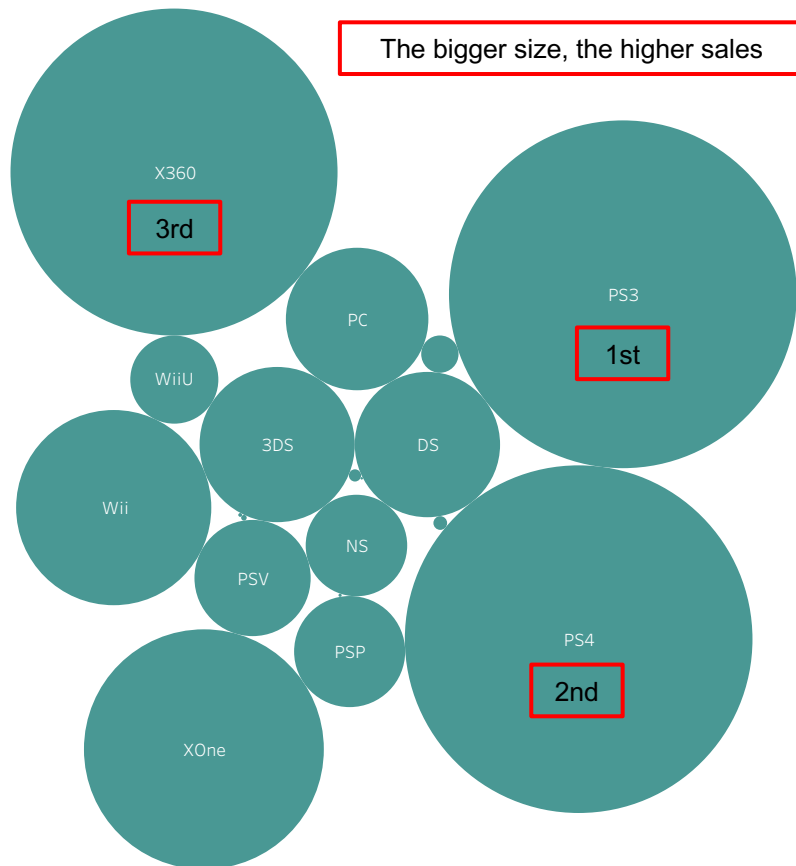
Platform. Size shows sum of Global Sales. The marks are labeled by Platform. The data is filtered on Year, which ranges from 2010 to 2020. The view is filtered on sum of Global Sales, which keeps non-Null values only.
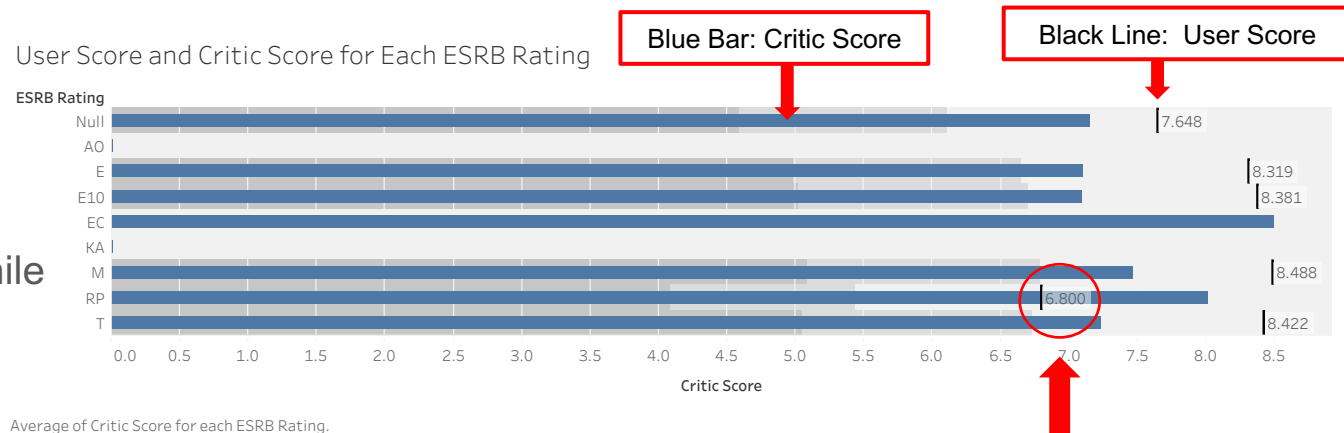
Visualization 8

Question 6
Is there any difference between user score and critic score within each ESRB rating?

Brief description
This visualization uses bullet graph to show both the user score and critic score in each ESRB rating. The blue bar represents the critic score, while the black line means the user score.

Key Takeaway
- For most ESRB Rating, user score shows higher value than critic score.
- In RP games, User score is obviously lower than critic score

User Score and Critic Score for Each ESRB Rating

Blue Bar: Critic Score

Black Line: User Score

7.648

8.319

8.381

8.488

6.800

8.422

Average of Critic Score for each ESRB Rating.

In ESRB Rating of RP, User score is lower than critic score

Reflection

In this assignment, I not only learn a lot of skill in Tableau, but also has more sense of designing chart which can convey the key point clearly. However, I still encountered some challenges during the process.

At first, I found that it is hard to find a good representation for ranking. Typically, lower ranking numbers should indicate higher sales, but using the rank itself to create a chart can potentially confuse viewers. Therefore, the solution is to use sales number directly. By displaying the actual sales values, viewers can still understand the ranking intuitively.

Another problem I faced is the missing data. Initially, I intended to identify the top 10 highest-selling games and their corresponding user scores. However, I discovered that some null values in the user score column made this task impossible to complete accurately. Excluding these data points would alter the composition of the top 10 games, thereby affecting the analysis results. Finally, I decided to focus on analyzing user scores separately in another aspect of the analysis, rather than incorporating them into the top 10 sales analysis.

Although it is challenging to do this assignment on a software I'm not familiar with at all, I overcome it by watching many videos that guided me through the implementation of Tableau. I really enjoy the moment when I find something new on my analysis. And I'm surprised to see that all of the initial questions have been answered successfully in this homework after making effort on it.