**Abstract - This paper will focus on information visualization tools to analyze the data of video game sales and find what factors contribute to video game sales.**

***Index Terms* --- Video Game, Nintendo, Activision, Ubisoft, Sony, Tableau, Orange, Rating, ESRB, RAWGraphs.**

1. **Introduction**

Video games are popular in these years whether adults or children. According to the new research released in 2019 by the Entertainment Software Association, over 164 million adults play video games in the United States. [1] There are many different kinds of games, and the games supported by each platform are also very different. So the question is, what kind of games are the most popular and what platforms sell the best. We will use Tableau, Orange, and other information visualization tools to analyze this dataset.

Data description is an important part of data visualization. The dataset was generated by a scrape of vgchartz.com by Gregory Smith and set simply extends the number of variables with another web scrape from Metacritic. [2] However, because meta critics cover only a subset of the platform, there is a lack of observations. In addition, the game may not have observations for all of the other variables discussed below. The total number of cases is 6900.

The entities from our dataset include *Name, Platform, Year\_of\_Release, Genre, Publisher, NA\_Sales, EU\_Sales, JP\_Sales, Other\_Sales, Global\_Sales*. In addition, there have any other entities which are rating for video games:

* *Critic\_score* - Aggregate score compiled by Metacritic staff
* *Critic\_count -* The number of critics used in coming up with the *Critic\_score*
* *User\_score* - Score by Metacritic's subscribers
* *User\_count -* Number of users who gave the *User\_score*
* *Developer* - Party responsible for creating the game
* *Rating* - The ESRB ratings(E.g. Everyone, Teen, Adults Only..etc) [3]

The next part presents what visualizations tools we use in this study.

1. **Visualization tools**

**2.1 Tableau Software**

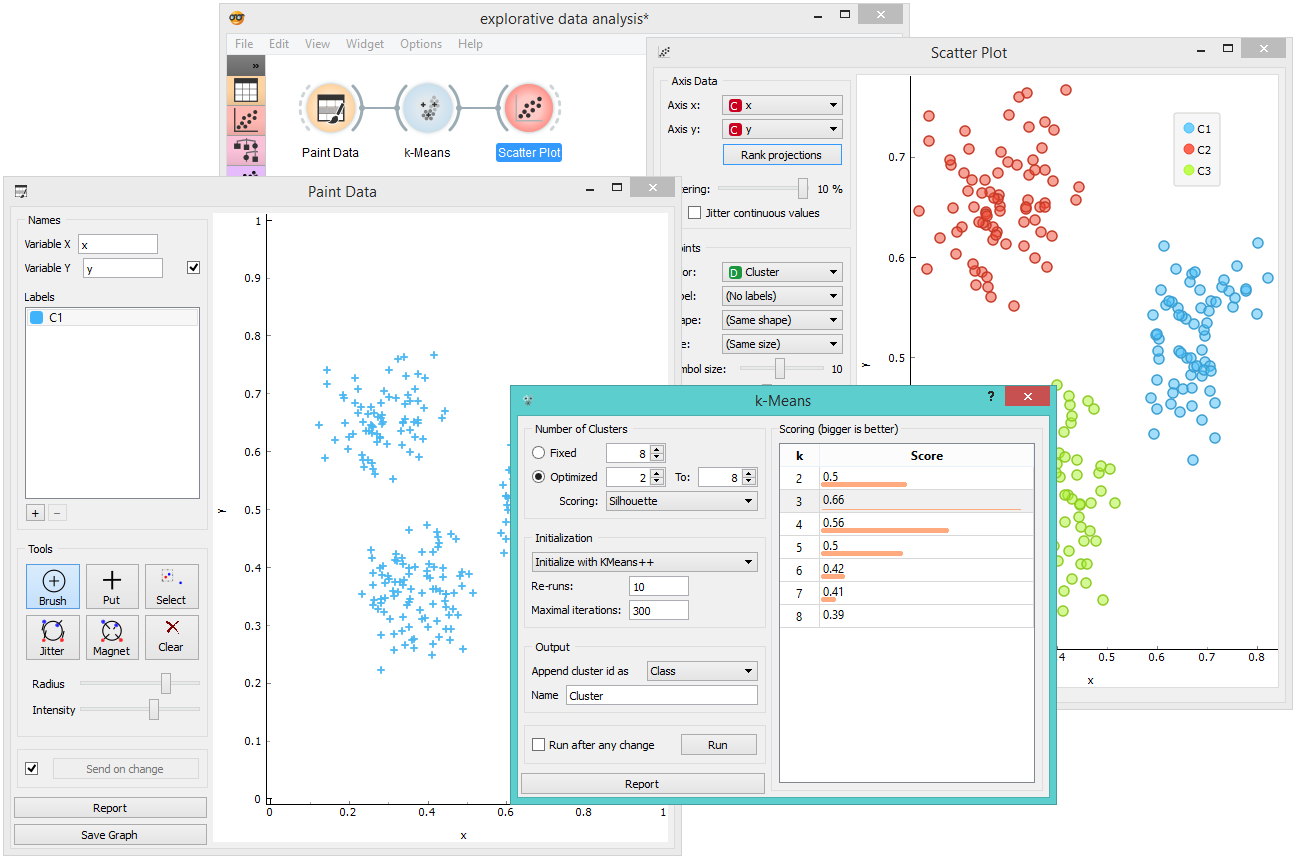
Tableau is a data analysis software, which is very simple to use. Through data import and data operation, it can analyze the data and generate visual charts to show the information people want to see. Tableau is fast, flexible and elegant to handle massive data. The greatest characteristic of Tableau is its simplicity and ease of use. Tableau users do not need to be proficient in complex programming and statistical principles. They just need to drag and drop the data directly into a workbook, and through some simple Settings, they can get the data visualization graphics they want. This means that we no longer need lots of teams of engineers, lots of time, custom software, and stale reports, and everyone can analyze the data as a service. In this report, Tableau can lead a more efficient way to do research and we treat it as our main visualization tool. Tableau is a good method to redesign existing visualization.For instance, adding more keys to generate visualization that will be better to analyze the difference of sales data between different genres or the relationship between rating scores and sales.

** Figure** **1** - Tableau

**2.2 Orange**

Orange is an open source machine learning and data visualization tool for novices and experts with many interactive data analysis workflows for data mining or machine learning models. In addition, it is tied to the Python language for scripting development. Components that comprise a range of data mining processes, such as data preprocessing, modeling, model evaluation, and visualization. [7]

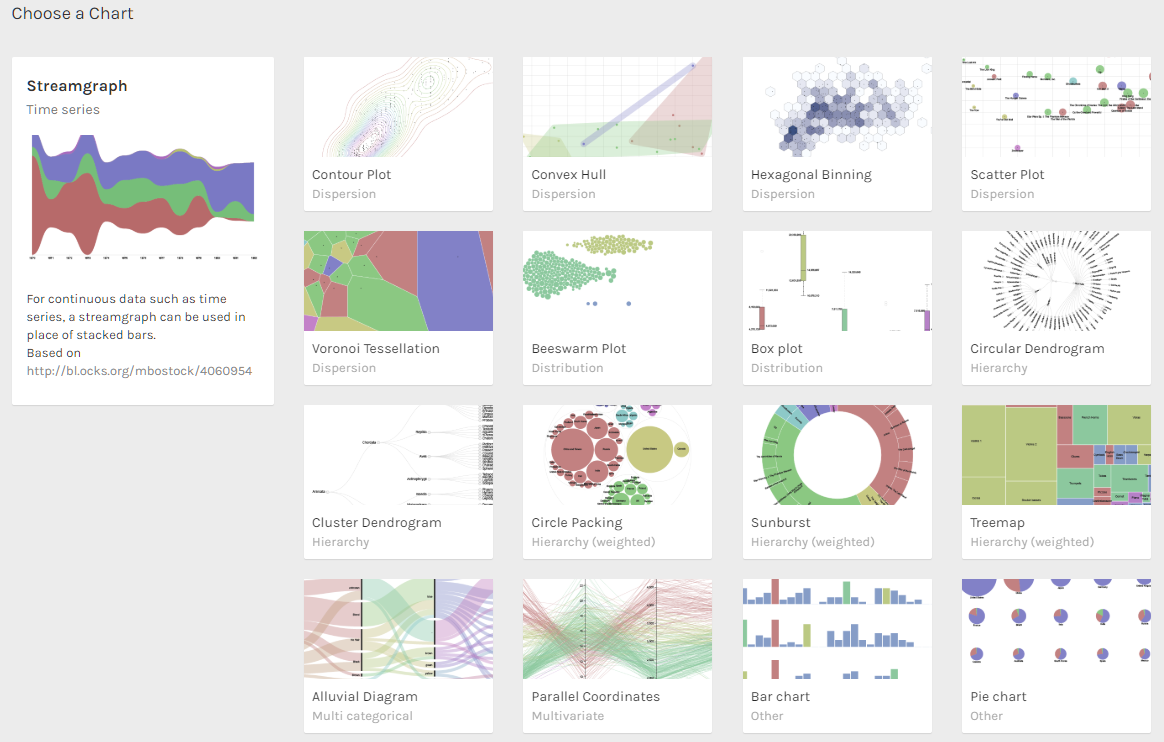
For this study, Orange can help to model the existing data and visualize it. We also use Orange to generate many types of visualizations, such as the difference of rating scores between different kinds of users. Furthermore, we use orange and Tableau to compare the results of visualizations. In this way, we could decide which visualization is better to show the correlation or other results, especially for the same research objects.

** Figure** **2** - Orange

**2.3 RAWGraphs.io**

RAWGraphs is an open source data visualization framework. RAWGraphs is intended to provide a missing link between the spreadsheet application and the vector graphics editor. You understand the need to process sensitive information, ensure that the data is secure, and that data uploaded to RAW is handled only by a Web browser.

The RAWGraphs is an easy way that leads us to gain an overall concept for our project. For instance, our group uses RAWGraphs to divide the whole dataset into some reasonable parts based on the genre which will help to finish later research. At the same time, RAWGraphs just provides us with the first step to be familiar with the dataset.

 **Figure** **3** - RAWGraphs

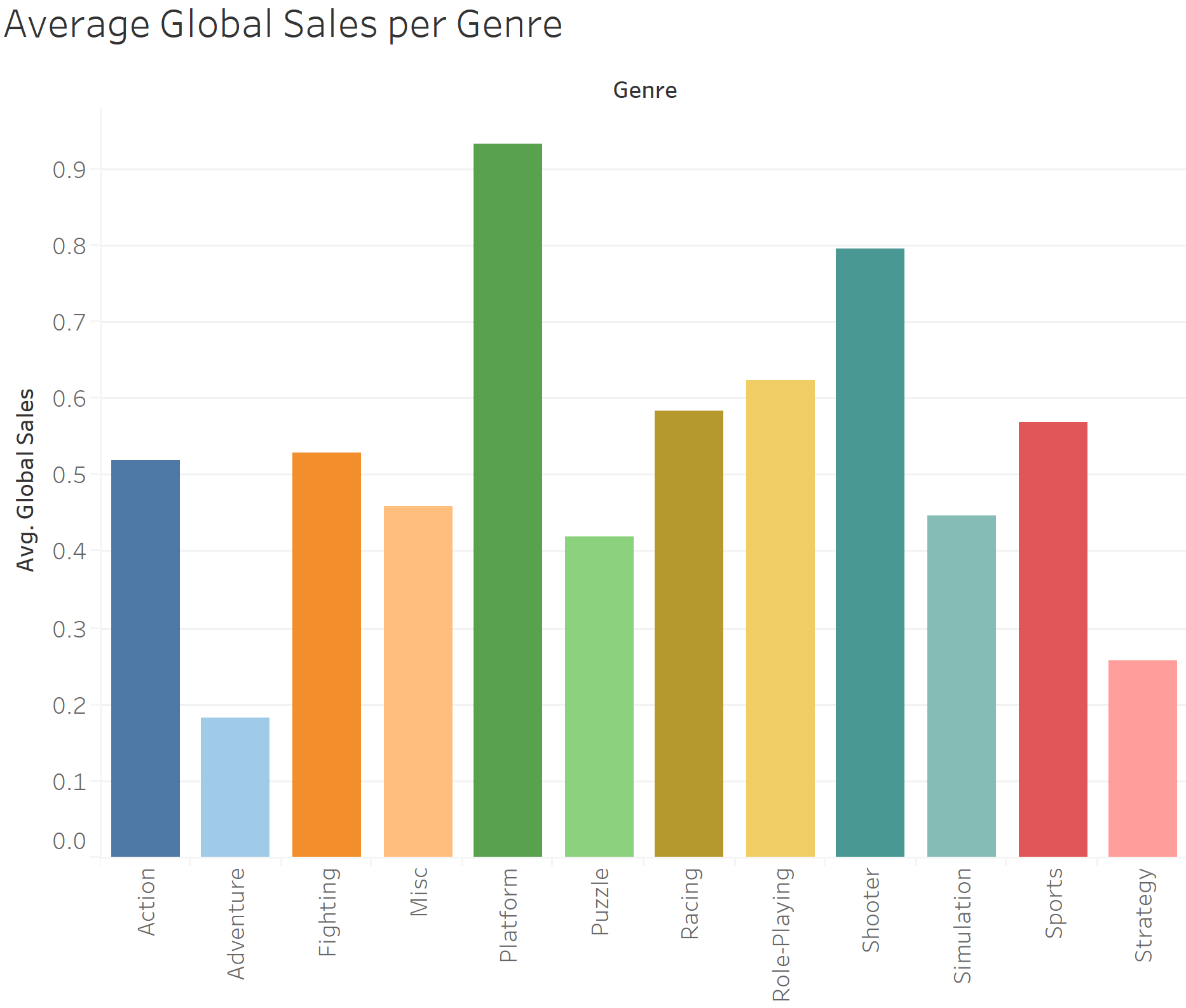
1. **Visualization Results**

**3.1 Video Game Sales**

Because we had tabular data, we decided to start with Tableau and RAWGraphs. First, we used RAWGraphs to visualize the global sales of games in each genre to get a sense of what composed our data set.

We decided that an easy way to visualize this was a multi-colored treemap. We can see from the treemap that action makes up the largest amount of global sales. However, we noticed that RAWGraphs does not allow us to add a key, so this graph is hard to understand unless you are told that size is based on global sales. For this reason, we decided to move to Tableau for the future visualizations.

**Figure 4** - Treemap Diagram

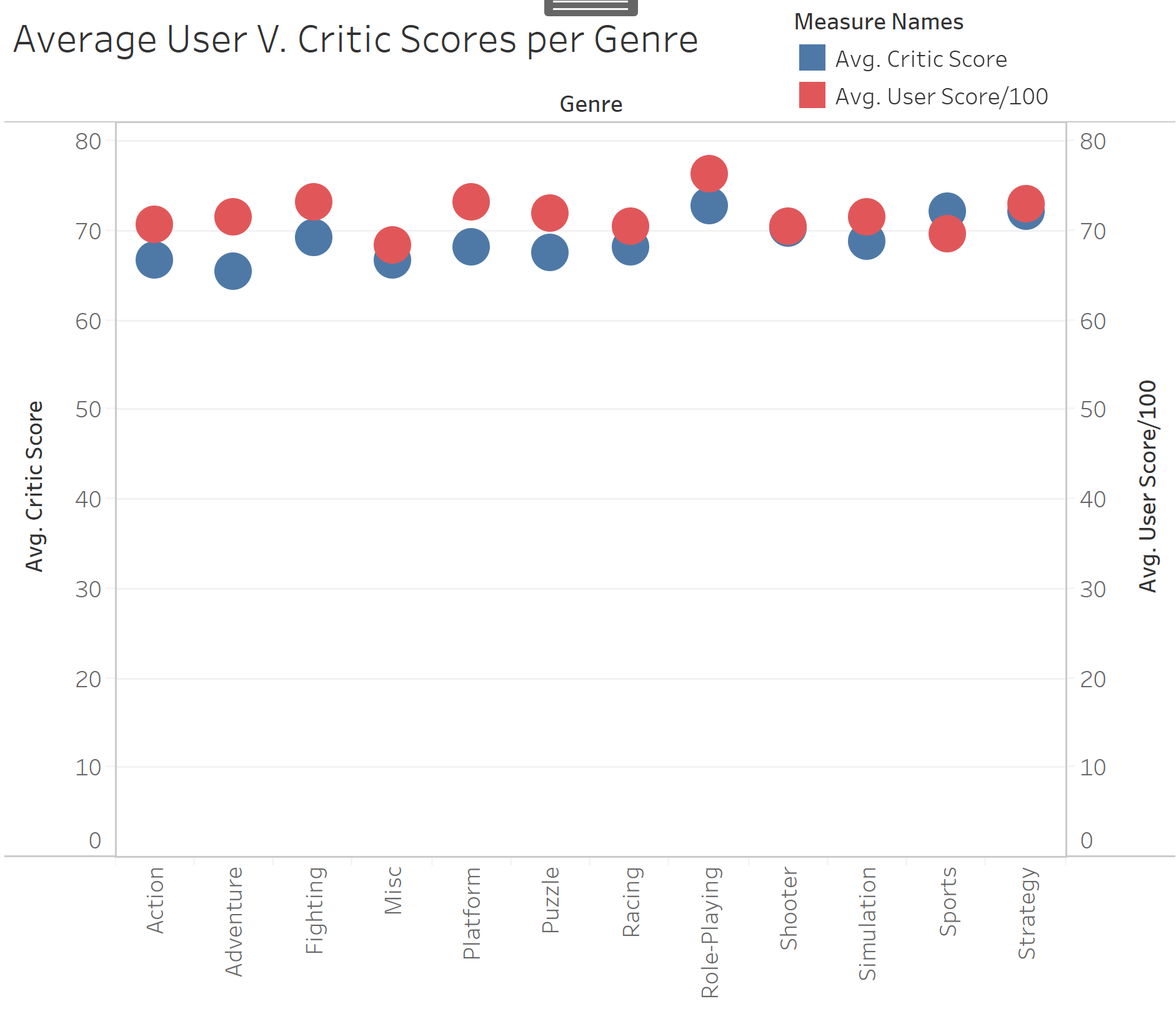
****

**Figure 5** - Bar Chart

Figure 5 is a redesign of the treemap above. We decided that, because there was no key on the treemap, we needed a better visualization depicting global sales per genre.

We can still see that action is, by far, the leader in global sales, but now we can better estimate which genres are close behind.

**3.2 Video Game Scores by Genre**

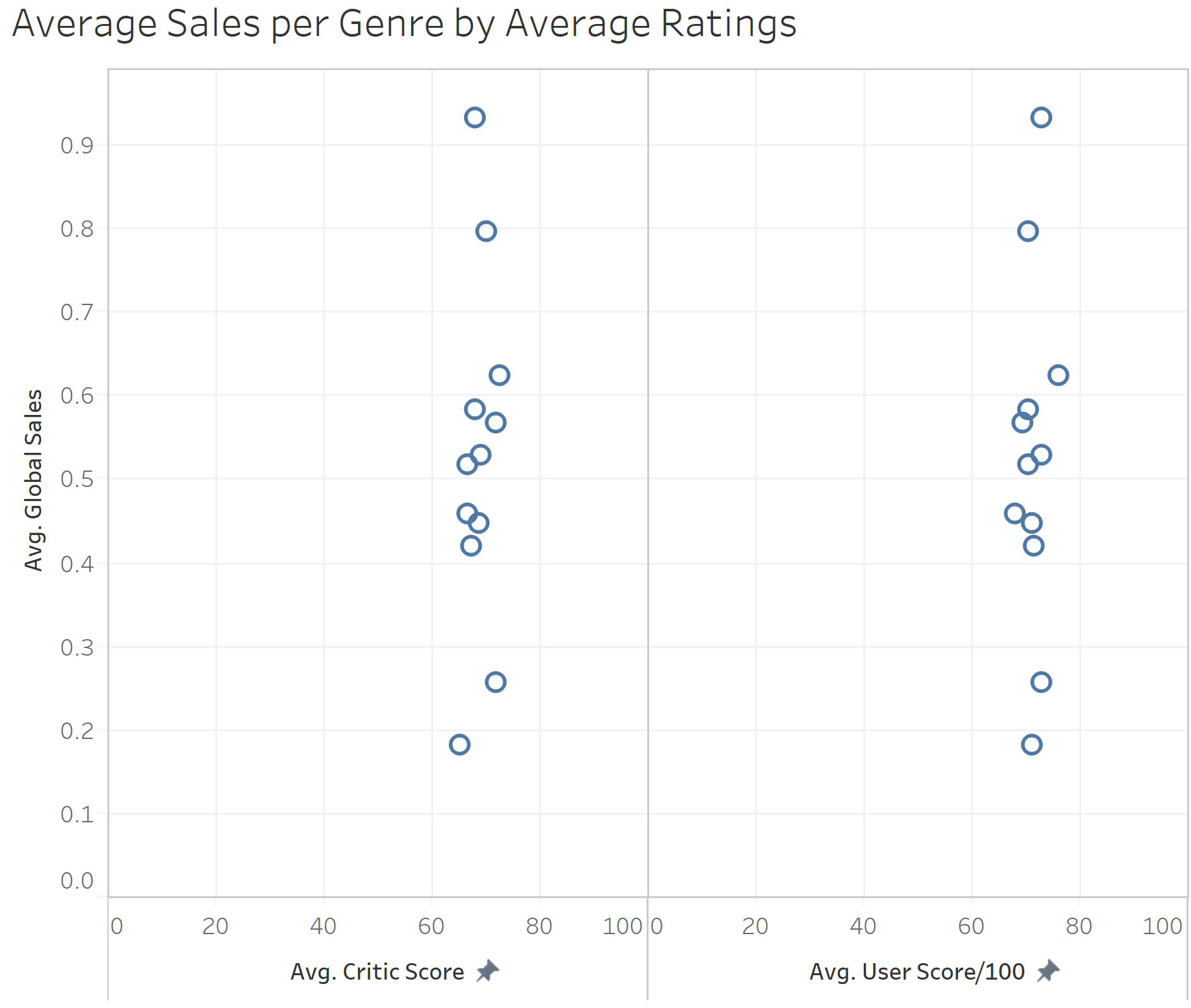
****

**Figure 6** - Scatterplot

Next, we used Tableau and started dragging and dropping fields in to see what types of correlations we could make. The first correlation that we found was user Scores versus critic scores per genre (figure 6). This is the average rating per genre that critics gave, versus what users gave. Because the user score was out of 10, while critic score was out of 100, we made a new calculated field that multiplied the user score by 10 to allow for a synchronized dual axis.

We found that the clearest way to visualize this comparison was using a scatter plot. We tried bar and stacked bar charts, but we found that the graph became too condensed, making it hard to read. Using a scatter plot with a synchronized, dual axis allowed us to easily compare one row based on color and height. This made it easy to see that, on average, users gave higher scores than critics did, for most genres.

**3.3 Average Sales per Genre by Average Ratings**

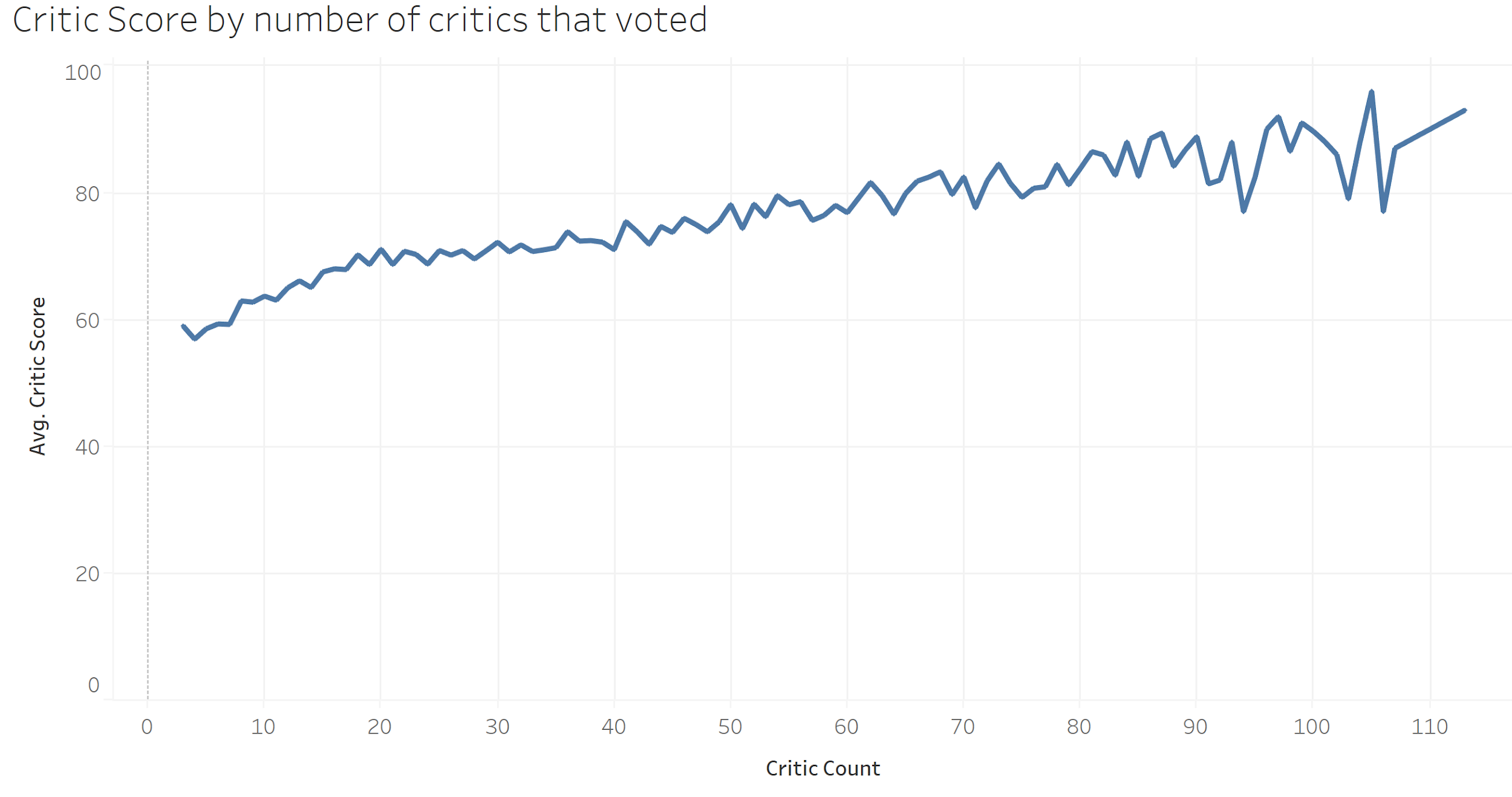
****

**Figure 7** - Bubble Chart

Next, since we had investigated both the average global sales and the scores by genre, we wanted to see if average sales were at all dependent on scores, and if so, were they dependent on critic scores or user scores. We thought that better user and critic scores would make people want to buy the game more, and therefore, games with higher critic scores would be higher in global sales.

For Figure 7, we made a scatter plot of average global sales to average critic scores, where each point is a different genre. We used genre because it was the dimension that we had used in our other visualizations. Our visualization showed that average global sales were not dependent on average rating at all, and in fact, unlike how genre impacted average sales to a great extent.

**3.4 Critic Score by Number of Critics that Voted**



**Figure 8** - Line Chart

Lastly, we decided to see if there was a correlation between how many critics voted for a game, and how highly that game was rated by critics. Figure 8 shows a line graph of this relationship.

We found that, surprisingly, there is a strong correlation between how many critics voted and average critic scores. We think that the more popular a game is, the more critics want to rate it, therefore, the better games are voted on more, and scored higher, but we were surprised to see how strong of a correlation was there.

1. **Conclusion**

Every year, there are numerous games being published and sold in the market. By 2020, The video games market is expected to be worth over 90 billion U.S. dollars, from nearly 78.61 billion in 2017.[4] It is an interesting and a meaningful thing to dig up the facts behind these sales. By using multiple visualization tools, we are able to understand the raw game sales data and look at them from totally new perspectives. Our analysis of Video Game Sales found many useful information and gave us some inspiration for our future work.

Critics and normal users are two main sources for video game reviews and scores. The differences between them are huge: normal users judge games casually, follow their feelings, while the critics give their reviews and scores based on industry standards because they have to follow some procedures in order to make professional comments and feedback, to make a living. However, normally both of these two groups give similar or even the same scores because the quality of the game is fixed during a certain timespan. Although average users may don’t know anything about game instudy standards, they can still feel the quality of a game. Our visualization(Figure 5) also shows this fact as we expected. For all game genres, the average user and critic scores are relatively close. For example, if the average user score of a game genre is around 70, the average critic score must be around to 70 as well. Another interesting fact is critic scores are normally higher than the user scores, but there is an exception which is the sports genre. We think the reason for this phenomenon is the particularity of the sports game. For other games, the game mechanics can be innovative that may make critics have more focus and commendations on them instead of other regular parts such as graphics, designs, etc. However, For sports games, the game mechanics are fixed since they are the derivatives of real life activities. The critics therefore become fastidious about the details such as graphics, etc. While normal users are satisfied as long as the game is playable, which causes this situation.

During the evolution of game development, many games with similar mechanisms are grouped to a certain kind of genre, and the number of these groups are constantly increasing and being studied: “Admitting genre evolution requires us to properly address the topic of innovation that lies at its basis. Consequently, a model of innovation specific to the video game industry has to be devised. Such a model will reveal that, far from being reducible to a simple checklist of specific game mechanics, video game genres play the part of the middle-man in a complex ecosystem of functional considerations and aesthetic ideas.”[5] By visualizing the raw data of “Video Game Sales with Ratings”[2](Figure 6), we are able to see which game genre is the most popular one among players while which one is relatively unpopular. According to our visualization, platform genre is the top seller in average global sales. It has at least 0.1 advantage compared to the second: shooter. The adventure genre is the most unpopular game genre in average global sales, it has about a 0.8 gap to the next one: strategy genre.

According to Figure 7, we also concluded the fact that the global sales of any game genre are neither affected by critic score nor user score. It is an inspirational fact because we thought customers’ purchase intentions can be influenced by other gamers’ reviews or scores. However, in fact, all game genres have their own long-term and increasing customer basis who are willing to purchase the product without prejudgment.

During our exploration, we also found “an empirical test of the "bandwagon effect" -- individuals rallying to the majority opinion.”[6] Based on our visualization(Figure 8), as the number of critic count increases, the average game score from critics increases as well. Although there are some throughs, the overall trend is still increasing. This means if a game gets more people’s attention, it will be more popular among gamers, so is its score.

Our next step can be to continue studying if this empirical “bandwagon effect” phenomenon exists in the user's score system as well. However, from this study, we not only learned many facts we didn’t know before from this dataset, but also found some proofs that linked to the knowledge we already know. We should use this experience and apply these visualization methods to other datasets in the future in order to gain insights, facts and practices.

1. **References**
2. Theesa. “2019 Essential Facts About the Computer and Video Game Industry”. Retrieved June 3, 2020 from: https://www.theesa.com/esa-research/2019-essential-facts-about-the-computer-and-video-game-industry/
3. Rush Kirubi. “Video Game Sales with Ratings: Video game sales from Vgchartz and corresponding ratings from Metacritic”. Retrieved June 3, 2020, from: https://www.kaggle.com/rush4ratio/video-game-sales-with-ratings
4. Entertainment Software Association(2020). “ESRB ratings” Retrieved June 3, 2020, from: https://www.esrb.org/
5. Wepc. “2020 Video Game Industry Statistics, Trends & Data”. Retrieved June 3, 2020 from: https://www.wepc.com/news/video-game-statistics/#video-gaming-industry-overview
6. Arsenault Eludamos. “Video Game Genre, Evolution and Innovation Dominic”. Journal for Computer Game Culture. 2009; 3 (2), p. 149-176. Retrieved June 3, 2020.
7. Richard Nadeau, Edouard Cloutier and J.-H. Guay. “New Evidence about the Existence of a Bandwagon Effect in the Opinion Formation Process”. International Political Science Review / Revue internationale de science politique. Vol. 14, No. 2 (Apr., 1993), pp. 203-213 (11 pages). Retrieved June 3, 2020.
8. Orange. “Data Mining Fruitful and Fun”. Retrieved June 3, 2020 from: https://orange.biolab.si/