Multivariate Analysis in Video Games sales *

Álvaro Novillo *Universidad Carlos III*
Polo *Universidad Carlos III*

In this article, we perform several dimensionality reduction techniques and clustering algorithms on a video game sales dataset available on Kaggle (https://www.kaggle.com/datasets/gregorut/videogamesales/data). Specifically, we use Principal Component Analysis (PCA) and Multidimensional Scaling (MDS) to reduce the dimensionality of the dataset. The article discusses the advantages and limitations of each technique and provides insights into the video game market based on the analysis.

Keywords: PCA, Videogames, Sales

About the dataset

The dataset under consideration contains information on video games with sales greater than 100,000 copies between 1980 and 2016. The dataset includes 11,493 unique game sales, detailing the name, year of release, genre, platform, and sales figures across numerous regions.

The dataset contains the following fields:

- **Rank** - Ranked by overall sales
- **Name** - Name of each videogame
- **Platform** - The games platform
- **Year** - Year of Release
- **Genre** - Genre of Game
- **Publisher** - Publisher of Game
- **NA_Sales** - Sales in NA (per Million)
- **EU_Sales** - Sales in EU (per Million)
- **JP_Sales** - Sales in JP (per Million)
- **Other_Sales** - Sales in ROW[1] (per Million)
- **Global_Sales** - Total worldwide sales (per Million)

Data Preprocessing

The dataset contains 11 variables, including quantitative variables like sales figures across various regions (NA_Sales, EU_Sales, JP_Sales, Other_Sales, and Global_Sales), the release year, and the rank of the game based on overall sales. Additionally, it includes multi-state categorical variables

*Replication files are available on the author's Github account (https://github.com/AlvaroNovillo). **Current version**: noviembre 07, 2023; **Corresponding author**: alvanovi@ucm.es.

[1]Net Sales (ROW) means the gross amount billed or invoiced on sales by Company and its Affiliates and Sublicensees of Licensed Products, less the following: (a) customary trade, quantity, or cash discounts and commissions to non-affiliated brokers or agents to the extent actually allowed and taken; (b) amounts repaid or credited by reason of rejection or return; (c) to the extent separately stated on purchase orders, invoices, or other documents of sale, any taxes or other governmental charges levied on the production, sale, transportation, delivery, or use of a Licensed Product which is paid by or on behalf of Company; (d) outbound transportation costs prepaid or allowed and costs of insurance in transit; and (e) allowance for bad debt that is customary and reasonable for the industry and in accordance with generally accepted accounting principles. ("Net Sales (ROW) Definition," n.d.)

like the genre, platform, and publisher of the game. To conform with the desired format, which requires at least two binary variables, we will filter out the video games of recent years and focus on titles that we are already acquainted with. Moreover, we will limit our research to two primary platforms, namely, Xbox One and PS4.

Table 1: Top five videogames, according to the sales ranking, that we are going to work with

	Rank	Name	Platform	Year	Genre	Publisher	NA_Sales	EU_Sales	JP_Sales	Other_Sales	Global_Sales
34	34	Call of Duty: Black Ops 3	PS4	2015	Shooter	Activision	5.77	5.81	0.35	2.31	14.24
78	78	FIFA 16	PS4	2015	Sports	Electronic Arts	1.11	6.06	0.06	1.26	8.49
93	93	Star Wars Battlefront (2015)	PS4	2015	Shooter	Electronic Arts	2.93	3.29	0.22	1.23	7.67
102	102	Call of Duty: Black Ops 3	XOne	2015	Shooter	Activision	4.52	2.09	0.01	0.67	7.30
110	110	Fallout 4	PS4	2015	Role-Playing	Bethesda Softworks	2.47	3.15	0.24	1.10	6.96
222	222	FIFA 17	PS4	2016	Sports	Electronic Arts	0.28	3.75	0.06	0.69	4.77

In Table 1. the top five selling games for 2015 and 2016, in PS4 ans Xbox One are shown. As we can see, the first one, which is Call Of Duty: Black Ops 3 is among the top 50 best selling games of the dataset (in PS4).

Examining the distribution of the filtered games rank, as seen in Figure @ref(fig:rank_distrib), considering its skewness, it can be confirmed that the vast majority of games released during this time period did not have a significant impact on the industry. In actuality, the average ranking of games within our dataset stands at 9373.

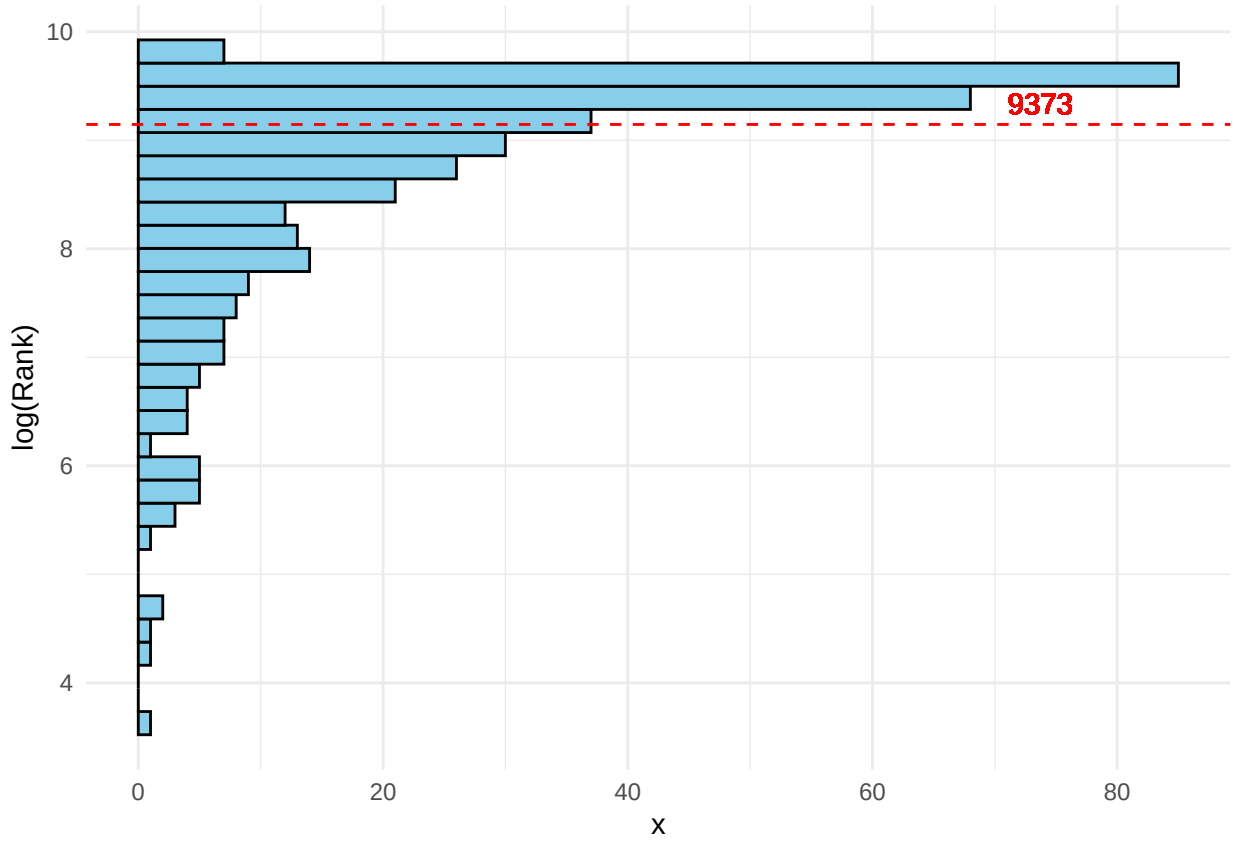

Figure 1: Distribution of the log-transformed Rank values. The red dashed line represents the median of the distribution

Figures **??** and **??** allow us to explore the basic features of our dataset, informing us of the amount of games from each platform, and the amount of games of each genre. In our dataset, the mayority of the sold games are from PS4, and the most popular genre is Action, followed by Sports, Role-Playing and Shooter

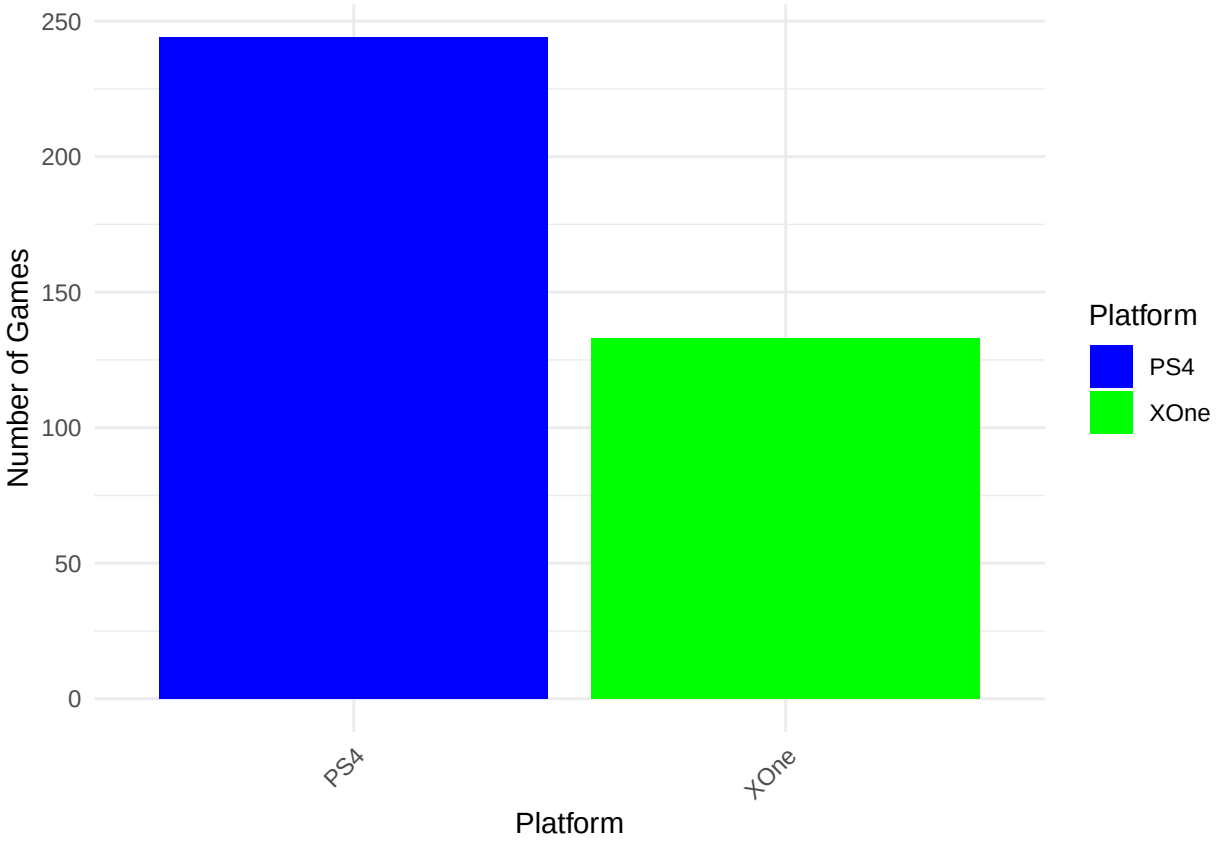

Figure 2: Number of games of each platform inside the dataset

Since the ranking is solely determined by overall sales figures, it is worthwhile investigating whether the top-selling game in certain regions differs from that of others. Our expectation is that the best-selling games in Japan will differ from those sold in the West.

The correlation between the sales in Japan and the West is 0.394, with an even lower correlation of 0.299 with the North American market, as illustrated in Figure **??**. This raises compelling questions about the underlying factors that contribute to these correlations. It is clear that several key factors highlight the significant differences between the Oriental and Western video game industries, leading to this low correlation.

First and foremost, the contrast in gaming preferences between regions plays a key role. In the West, specifically in North America, action and shooter games are incredibly popular. However, the Japanese market favours Role Playing Games (RPGs), which differs greatly from the Western market. As a result of these diverging gaming genres, differing sales patterns naturally occur, ultimately contributing to the observed low correlation.

The marketing and localization strategies utilised in the Japanese video game industry are of great significance. Numerous Japanese games are designed with a primary focus on the local market, giving rise to gameplay and cultural elements that may not resonate as strongly with

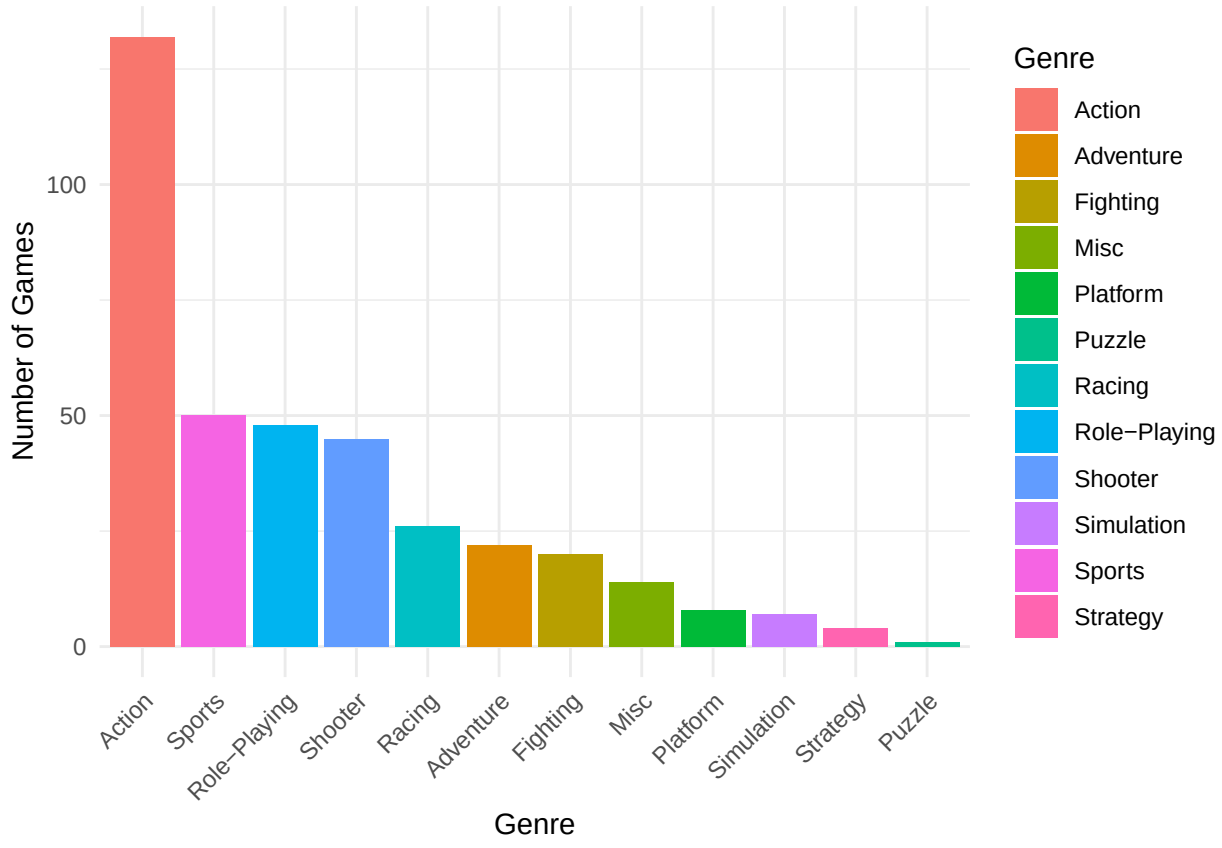

Figure 3: Barplot of the amount of games of each genre

Western or North American audiences. Therefore, these games may not be successful beyond their intended audience in the East, resulting in a larger sales gap and a weaker association with these regions.

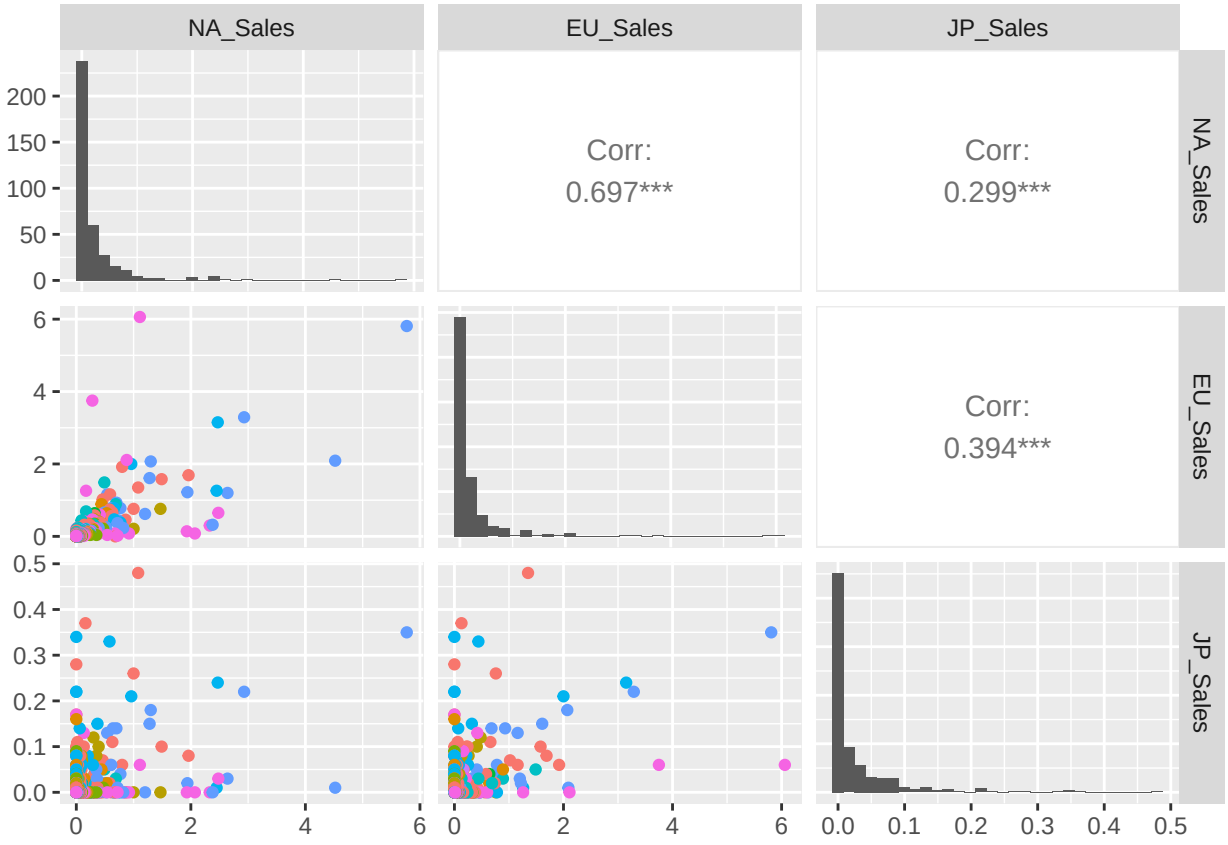

Figure 4: Correlation Plot of Sales Variables

To delve deeper into that matter, Table **??** presents a comprehensive analysis of the percentage distribution of sales across the top three genres within diverse regions under investigation. It is evident from the table that the genre of Role-Playing Games (RPGs) enjoys significantly greater popularity in Japan as compared to North America and Europe. Strikingly, our research reveals that Action games emerge as the most prevalent genre in Japan, accounting for a substantial portion of the region's total sales, encompassing 35.28% of the market share.

Table 2: Percentage distribution of sales for the top three genres in different regions

Genre	Percentage_NA_Sales	Percentage_EU_Sales	Percentage_JP_Sales
Action	20.79	23.19	35.28
Role-Playing	11.75	11.81	27.26
Shooter	35.69	29.99	15.38

References

"Net Sales (ROW) Definition." n.d. https://www.lawinsider.com/dictionary/net-sales-row.