Driving Economic Growth: The Impact of Video Game Sales (2006-2010)

June 16, 2024

Group 10:
1. Kevin Husodo (2602126896)
2. Putri Febiyani (2602181875)
3. Audric Nagata (2602090435)

1. Introduction

A. Purpose, Context, and Structure of the Report

This report aims to analyze video game sales data from 2006 to 2010, sourced from a dataset available on the Kaggle platform. Kaggle is renowned for its rich data sources frequently utilized by data analysts and researchers for various analytical projects. The report provides comprehensive insights into video game sales trends, genre popularity, regional sales performance, and an analysis of leading platforms and publishers during the specified period.

The structure of this report includes several main sections:

- **Introduction**: Outlining the objectives, context, and structure of the report.

- **Dataset Description**: Explaining the data source, data structure, and variables used in the analysis.

- **Exploratory Data Analysis (EDA)**: Presenting initial analysis results, including sales trends, genre popularity, and regional performance comparisons.

- **Key Insights**: Identifying the main findings from the data analysis.

- **Conclusion**: Summarizing key findings and providing recommendations for further actions.

B. Introduction to the Topic or Issue Addressed

The main topic of this report is the analysis of video game sales exceeding 100,000 copies from 2006 to 2010. Video game sales are a crucial indicator reflecting the popularity and success of games in the market.

The issues addressed in this report include:

- How did video game sales trends evolve from year to year during the period from 2006 to 2010?

- Which video game genres were most popular, and how did their popularity evolve during this period?

- How did video game sales perform in various geographic regions such as North America, Europe, Japan, and other regions?

- Which gaming platforms exhibited the best performance during the analysis period?

- Who were the publishers with the highest sales, and what were their successful games?

C. Significance of the Analysis

This analysis is significant as it provides a deep understanding of the dynamics of the gaming industry from 2006 to 2010. The video game industry is one of the most dynamic and rapidly evolving entertainment sectors. By understanding sales trends, genre popularity, and regional performance, game developers, publishers, and marketers can make more informed and strategic decisions. This analysis is valuable not only for understanding the current market conditions but also for predicting future trends and identifying new opportunities.

D. Relevance and Potential Implications

This analysis is relevant to various stakeholders interested in the dynamics of the gaming industry. The findings can be utilized for:

- **Game Development**: Identifying genres and platforms worthy of focus in game development.

- **Marketing Strategies**: Optimizing marketing strategies based on regional sales performance and market trends.

- **Investment Decisions**: Assisting investors in understanding market dynamics and identifying potential areas for investment.

- **Competitive Analysis**: Providing insights into successful titles and strategies of competitors in the gaming industry.

E. Target Audience

This report is intended for:

- **Researchers**: Interested in data analysis of the gaming industry and market trends.

- **Stakeholders in the Gaming Industry**: Including developers, publishers, and marketers needing insights to support their business decisions.

- **Investors**: Seeking to understand the dynamics of the gaming market for better investment decisions.

- **General Audience**: Anyone interested in insights into video game sales and the development of this industry.

F. Objectives and Analytical Questions

The main objectives of this analysis are to answer several key questions:

1. What were the video game sales trends from year to year during the period from 2006 to 2010?

2. Which video game genres were the most popular, and how did their popularity evolve during this period?

3. How did sales performance vary across different regions?

4. Who were the publishers with the highest sales, and what were their successful games?

5. How did various gaming platforms perform over time?

G. Methods and Analytical Techniques

This analysis employs Exploratory Data Analysis (EDA) methods, including:

- **Exploratory Data Analysis (EDA)**:
Is an approach to analyzing datasets with the aim of summarizing their main characteristics using visual methods. This technique includes data visualization using graphs such as histograms, scatter plots, and box plots to understand data distribution, relationships between variables, and trends over time. Statistical analysis is also applied to measure correlations between variables, convey the significance of findings, and test hypotheses to understand the data more deeply before it is formally modeled [1].

- **The Kruskal-Wallis**:
Is used as a non-parametric alternative to the ANOVA test, useful for comparing the medians of three or more independent groups of data. In the context of video game sales analysis, this test is used to see whether there are significant differences in sales between regions or certain game genres [2].

- **The Chi-Square test**:
  Is a statistical technique for testing the relationship between two categorical variables. In video game sales analysis, this test is used to determine whether there is a relationship between game genre preferences and market region, or whether there is a dependency between the popularity of a particular platform and game sales by genre [3].

H. Assumptions and Limitations

Some assumptions and limitations of this analysis include:

- **Data Accuracy**: The data used is sourced from Kaggle for the period 2006-2010. While Kaggle is known for rich datasets, it may not encompass all recent changes in the video game industry. For example, trends from 2017-2020 may not be well-reflected due to limited data availability for those years.

- **Time Constraints**: The analysis is constrained by the timeframe of data from 2006 to 2010. Significant changes in market preferences and technology in the gaming industry after 2010 may not be fully captured in this report.

- **Lack of Recent Data**: Despite efforts to incorporate newer data, availability for years 2017-2020 has been limited or inadequate in the available sources. This may affect the balance of representation in our visualizations and analysis.

- **Industry Dynamics**: The gaming industry is a highly dynamic market with rapid changes in technology, consumer preferences, and global trends. These factors can influence the analysis outcomes based on historical data used in this report.

- **Interpretation of Results**: The results of this analysis should be understood in the context of the timeframe and data used. Interpretations regarding genre preferences, platforms, and business strategies should be adjusted for contextual changes and current market dynamics, which may differ from the period of 2006-2010.

I. Personal Relevance

This analysis is worth reading as it provides actionable insights into business strategies within the gaming industry. For us, this analysis is compelling as it combines our interest in data and the gaming industry, offering an opportunity to apply data analysis skills in a real-world context. Understanding trends and preferences in the gaming industry can be crucial for success across various aspects of the video game business. This analysis also provides an opportunity to delve deeper into how specific factors impact game success, which can be highly valuable to anyone involved in this dynamic industry.

2. Data Description

A. Dataset Description

The dataset contains information on video game sales exceeding 100,000 copies. It was sourced from Kaggle, which originally obtained the data from vgchartz.com. The dataset spans from 1980 to 2020, but the analysis focuses on the years 2006 to 2010.

B. Variables Involved

1. Name:

- Description: Provides the name of each video game in the dataset.

- Significance: This variable serves as a unique identifier for each data entry, distinguishing between different games for analysis purposes.

2. Platform:

- Description: Indicates the platform on which each video game was released (e.g., PC, PS4, Xbox One).

- Significance: The release platform can influence the game's market accessibility. Analyzing platforms also offers insights into consumer preferences for specific gaming platforms.

3. Year:

- Description: Represents the year when each video game was released.

- Significance: Information on release year is crucial for understanding sales patterns over time. Changes in technology and consumer preferences can impact game sales performance annually.

4. Genre:

- Description: Specifies the category or genre of each video game (e.g., action, adventure, sports).

- Significance: Game genres provide insights into thematic preferences and popular game types among consumers. Genre analysis aids in understanding market trends and strategic game development *decisions.

5. Publisher:

- Description: Names the publisher or company that released the video game.
- Significance: Identifying the publisher is important as publishers often employ distinct marketing and distribution strategies that influence game sales

performance. Major publishers also wield significant impact on game popularity and distribution.

6. NA_Sales, EU_Sales, JP_Sales, Other_Sales, Global_Sales:

- Description: Each column represents video game sales in millions of copies in specific regions (North America, Europe, Japan, other regions worldwide, and total global sales).

- Significance: These variables offer a comprehensive view of the commercial performance of games across different regional and global markets. Understanding sales distribution across regions aids in adjusting marketing and distribution strategies to enhance revenue.

C. Summary Data

Rank	Name	Platform	Year	Genre	Publisher	NA_Sales	EU_Sales	JP_Sales	Other_Sales	Global_Sales
Min. : 1	Length :16598	Length :16598	Length :16598	Length :16598	Length :16598	Min. : 0.0000	Min. : 0.0000	Min. : 0.00000	Min. : 0.00000	Min. : 0.0100
1st Qu.: 4151	Class :character	Class :character	Class :character	Class :character	Class :character	1st Qu.: 0.0000	1st Qu.: 0.0000	1st Qu.: 0.00000	1st Qu.: 0.00000	1st Qu.: 0.0600
Median : 8300	Mode :character	Mode :character	Mode :character	Mode :character	Mode :character	Median : 0.0800	Median : 0.0200	Median : 0.00000	Median : 0.01000	Median : 0.1700
Mean : 8301						Mean : 0.2647	Mean : 0.1467	Mean : 0.07778	Mean : 0.04806	Mean : 0.5374
3rd Qu.:12450						3rd Qu.: 0.2400	3rd Qu.: 0.1100	3rd Qu.: 0.04000	3rd Qu.: 0.04000	3rd Qu.: 0.4700
Max. :16600						Max. :41.4900	Max. :29.0200	Max. :10.22000	Max. :10.57000	Max. :82.7400

The summary data provides a comprehensive overview of video game sales from a dataset comprising 16,598 entries. Key variables include rank, game name, platform, release year, genre, and publisher, all categorized as character data types. Sales figures across four major regions (North America, Europe, Japan, and other regions) are reported in millions of copies, with global sales also in million units. The rank ranges from 1 to 16,600, with median and mean values around 8,300, indicating a relatively even distribution. The significant difference between median regional and global sales, notably a global median of 0.17 million copies, highlights varied market preferences.

3. Data Preprocessing

Here are the steps of preprocessing applied to the dataset, along with their reasons

A. Prepare data

```r
# Visualizations
library(ggplot2)
library(corrplot)
library(reshape2)
library(tidyr)
library(knitr)
library(plotly)
library(RColorBrewer)
# Data Manipulation
library(dplyr)
```

Libraries Overview:

Library	Description
ggplot2	A core library for creating customizable data visualizations in R. It uses a layered graphics grammar to create plots step-by-step, offering flexibility.
corrplot	Used to visualize correlation matrices in R with various visualization methods.
reshape2	R library for restructuring data between wide and long formats, facilitating data analysis and visualization.
tidyr	Enables data transformations such as reformatting, cleaning, and preparing data for analysis, adhering to tidy data principles.
knitr	Essential for generating dynamic reports in R, supports Markdown and HTML-based documentation.
plotly	Enables creation of interactive web-based visualizations directly from R, allowing dynamic and shareable charts.
RColorBrewer	A package that provides a collection of well-designed color schemes for data visualization in R.
dplyr	Provides tools for efficient dataset manipulation in R, including filtering, summarizing, and transforming data, crucial for data preprocessing.

B. Handling Missing Data

```r
data <- dataVG[!is.na(dataVG$Year) & dataVG$Year != "N/A", ]
```

Reason: Removing rows where the Year column is NaN or "N/A" ensures that only complete and valid data are used. Missing data can introduce bias and reduce the reliability of models. By removing them, we ensure that the analysis is based on accurate information.

C. Filtering Records based on Specific Years

```r
data <- data[data$Year %in% c("2008", "2009", "2007", "2010", "2006"), ]
```

Reason: Selecting only rows with specific years (2006, 2007, 2008, 2009, 2010) focuses the analysis on a relevant time period or recent data that is more important for the current analysis. This reduces noise and increases the relevance of analysis results by discarding potentially irrelevant data.

D. Changing Year Column to a Factor

```r
data$Year <- factor(data$Year)
```

Reason: Changing the Year column to a factor is necessary because years are typically used as categorical variables in statistical analysis and visualization. By converting it to a factor, we can easily perform grouping operations, create contingency tables, and make relevant plots such as histograms or bar charts to understand the distribution of data across years.

By following these preprocessing steps, the data becomes cleaner, more focused, and ready for further analysis. Good preprocessing helps minimize potential errors or biases in interpreting analysis results and enhances the ability to derive meaningful insights from the available data.

4. Data Exploration

A. Boxplots of Sales Data


Boxplots of Sales Data
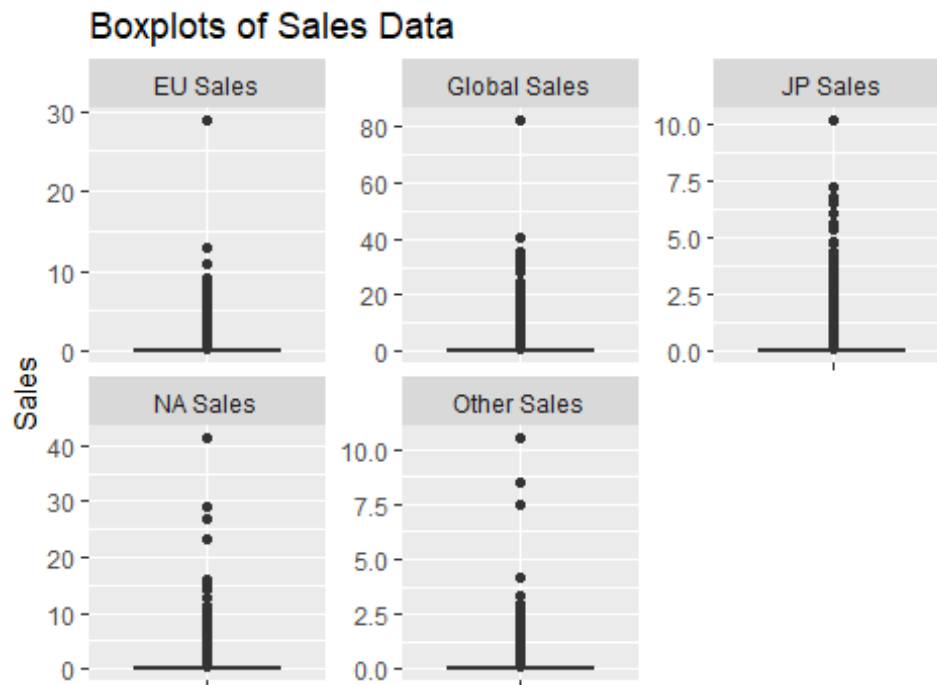
Key Observations

- **Outliers Across Regions**:
  The visualization shows a single outlier in each region.
- **Median Distribution**:
  All region's median score is almost similar showing a near-zero score.
- **Regional Variances**:
  While Action remains dominant, the gap between Action and other genres varies by region. It's notably larger in North America and Europe compared to Japan.
- **Interpretation**:
  The visualization shows that the sales of video games everywhere in the world is rather flat and distributed by the same manner.

Insights

- **Sales Strategy**: The variability and presence of outliers in every region indicate a need for a more tailored sales strategy. Identifying why certain games perform exceptionally well or poorly can help in developing better marketing and development plans.

- **Global Strategy**: Recognizing the game preference for everyone could be a strategy to release games in the future.
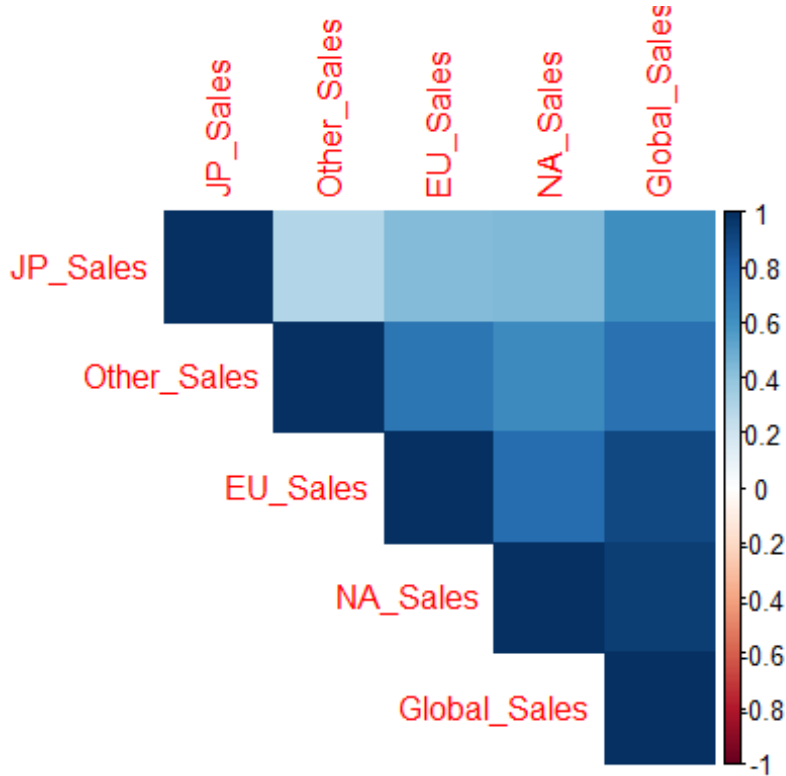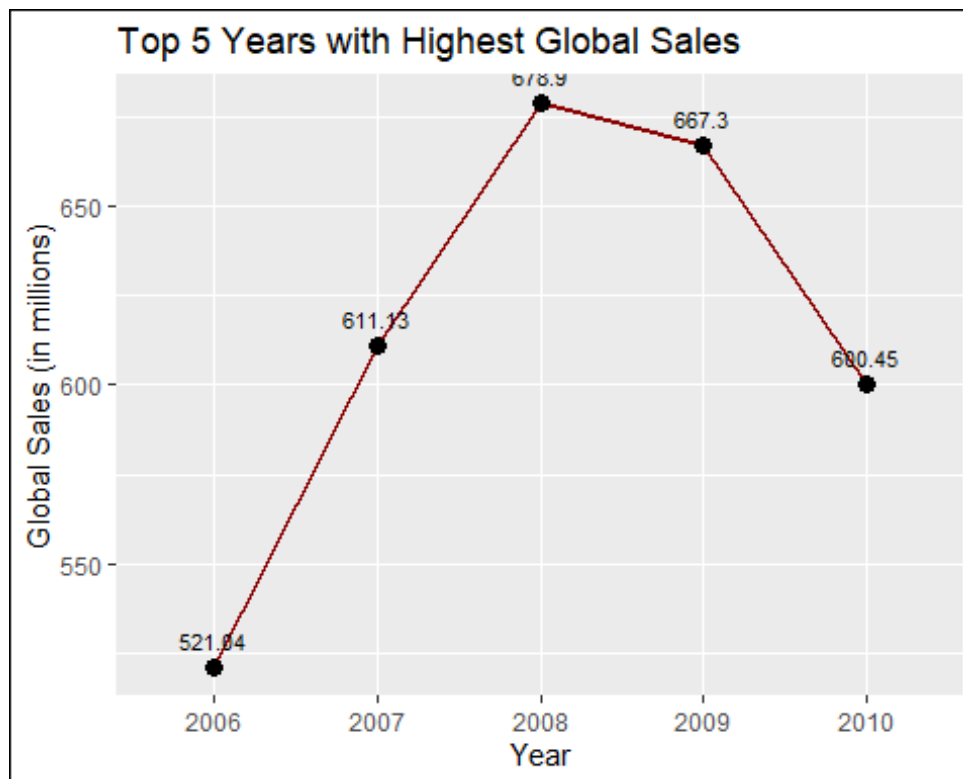
Key Observations

- **Contribution from each region towards global sales**:
  The visualization interprets the different sales across regions make up the global sales. Japan shows the least contribution and North America shows the most contribution.
- **Population influence**:
  North America sales implies that the more the population size is, the more will they contribute to global sales of video games. Larger populations mean more individuals are likely to purchase video games, spanning a diverse range of age groups and demographics. Additionally, higher population densities in urban areas facilitate better marketing reach and distribution efficiencies.

- **Interpretation**:
  Trends play a major role in sales, as seen in the visualization. The success of sales in Europe, Japan, and North America defines the video game trend in these three regions. In contrast, other regions, despite having larger populations, cannot compete with the video game sales of these regions.

Insights

- **High-Priority Markets**: Given the importance of Europe, Japan, and North America, game developers and marketers should prioritize these regions for product launches, marketing campaigns, and tailored content.

- **Localized Content**: Since these regions define trends, creating localized content that caters to the specific preferences of gamers in Europe, Japan, and North America can boost sales and engagement.

C. Top 5 Years with Highest Global Sales
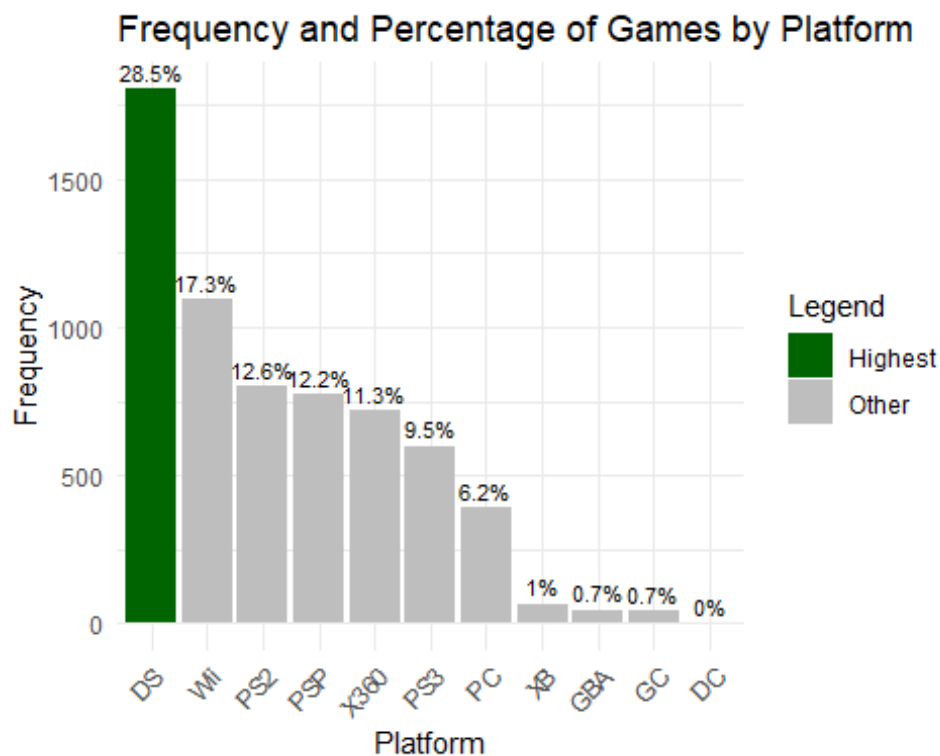

Key Observations

- **Sales peaks and troughs**:
  The visualization indicates that 2008 represents the highest peak of sales, while 2006 marks the lowest trough.

- **Irregular patterns**:
  The visualization reveals inconsistent sales trends before and after 2008. Specifically, there is a significant leap in sales from 2006 to 2007, with a notable increase of 100 million units. In contrast, after 2008, all sales go down slowly.

- **Interpretation**:
  The upward trend in the line graph suggests a growth trajectory for global sales over the past five years. The year with the possible decline might be an outlier or indicate a temporary setback.

Insights

- It is possible that the company is expanding into new markets, or that its products are becoming more popular. The increase in sales could also be due to a general improvement in the global economy.

D. Frequency and Percentage of Games by Platform

Key Observations

- **Platform Dominance**:
 DS is the most dominant platform, accounting for roughly 28.5% of the games. Wii follows closely at 17.3%.
- **Platform Distribution**:
  There is a variety of platforms represented, with a clear distinction between platforms released by Nintendo and other consoles. Platforms like PC, Playstation2, Playstation Portable (PSP), and Xbox 360 (X360) hold a smaller share compared to DS and Wii.

- **Interpretation**:
  The data suggests that Nintendo DS is the most popular platform, followed by a console produced also by Nintendo called Wii. A variety of other platforms exist but hold a smaller market share.

Insights

- This visualization highlights the dominance of Nintendo compared to other console gaming in terms of the number of games released. It's important to consider that this might not reflect sales figures, as some platforms might be more popular for specific genres or have a larger average number of games released per title.

E. Frequency and Percentage of Games by Genre
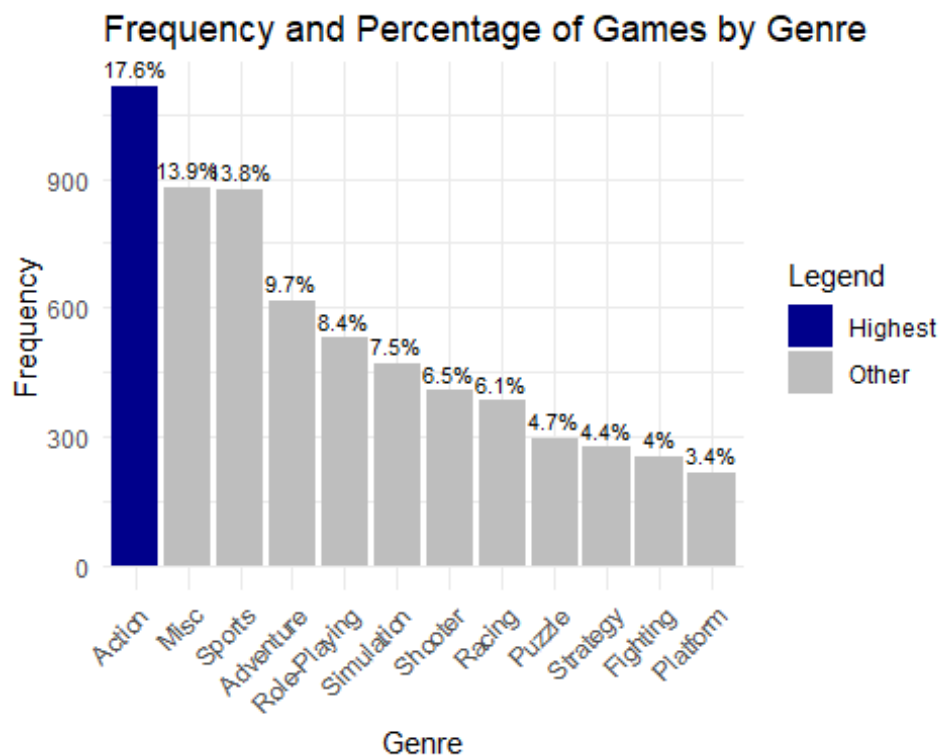
Key Observations

- **Genre Distribution**:
  Action games represent the highest frequency and percentage of games overall, followed by Misc and Sports genres.
- **Genre Percentages**:
  The percentage of games in each genre seems to be relatively stable across the years, with some minor fluctuations.

- **Interpretation**:
  Action games are consistently the most popular genre, while other genres appear to have a steady presence in the market. The introduction of the "Fighting" genre in

2017 might suggest a growing interest in this genre or a shift in categorization methods.

Insights

- This visualization suggests that the video game market offers a variety of genres to cater to different player preferences. Action games seem to be dominant, but other genres have a steady presence as well.

F. Top 10 Names Game by Sales in Different Regions

```
## PhantomJS not found. You can install it with webshot::install_phantomjs().
If it is installed, please make sure the phantomjs executable can be found
via the PATH variable.
```

Key Observations

- **Top Selling Games**:
  The chart shows the top 10 games by global sales, with Wii Sports leading the pack at nearly 80 million copies sold globally. Other titles from Nintendo's Wii console, including Mario Kart Wii and Wii Sports Resort, also rank high on the list.

- **Genre**:
  All the top 10 selling games appear to be casual or family-oriented games, possibly due to the broad appeal of the Wii console.

- **Interpretation**:
  The data suggests that during this timeframe, casual games on the Wii console were dominant in global sales. This could be due to the unique motion-control features of the Wii that made gaming more accessible to a wider audience.

Insights

- The data suggests that during this timeframe, casual games on the Wii console were dominant in global sales. This could be due to the unique motion-control features of the Wii that made gaming more accessible to a wider audience.

G. Platform Frequency by Year

Key Observations

- **Platform Popularity**:
  The graph shows the number of players across all platforms within the timeframe. The DS platform appears to be the most popular, with a steady increase in the number of players over the five years displayed (2006-2010).

- **Player Trends**:
  The number of DS, PS3, and Wii players appear to fluctuate throughout the timeframe, while the number of GBA players steadily declines over the five years.

- **Interpretation**:
  The data suggests that the DS platform gained significant popularity during the timeframe, whereas PS3 rising steadily and PS2 console users declined. This could be due to the portability and new features of the DS, like the touchscreen.

Insights

- This visualization highlights a potential shift in gaming preferences towards mobile platforms around the late 2000s. It would be insigtful to explore if the popularity of the DS translated to a rise in mobile gaming overall, or if it catered to a new audience to mobile gaming.

H. Top 10 Game Genres by Sales in Different Regions
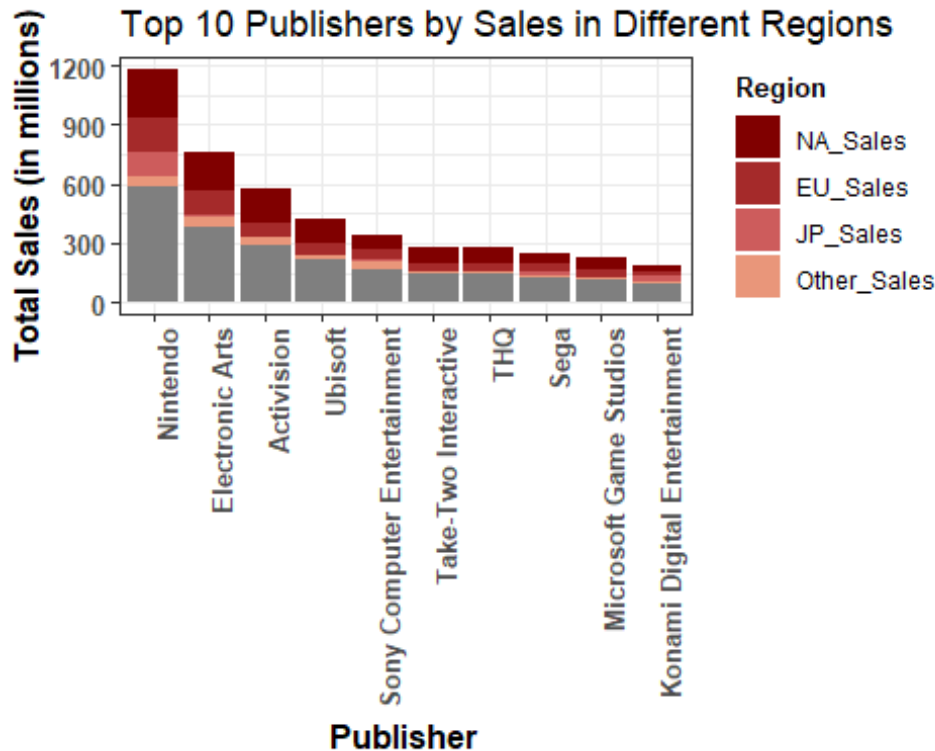
Key Observations

- **Genre Popularity Across Regions**:
  The visualization highlights how the popularity of video game genres varies across different regions: North America (NA), Europe (EU), Japan (JP), and other regions.

- **Dominance of Action Genre**:
  Across all regions, the Action genre emerges as the top seller, indicating its widespread appeal globally.
- **Regional Variances**: While Action remains dominant, the gap between Action and other genres varies by region. It's notably larger in North America and Europe compared to Japan.
- **Role-Playing Preference in Japan**: Role-Playing games show a stronger preference in Japan relative to other regions, suggesting cultural or market-specific influences.
- **Interpretation**:
  The visualization suggests that while certain genres like Action maintain universal popularity, regional preferences and market dynamics play a significant role in shaping genre sales distributions.

Insights

- **Market Segmentation**:
  Tailoring marketing and content strategies based on regional preferences can enhance market penetration and consumer engagement.

- **Global Strategy**:
  Recognizing the dominance of Action while capitalizing on niche preferences (Role-Playing in Japan) can inform strategic decisions for global game releases.

Top 10 Publishers by Sales in Different Regions

Key Observations

- **Top Publishers by Sales in Different Regions**:
  The chart displays the top publishers by total sales across four regions: North America (NA), Europe (EU), Japan (JP), and Other.

- **Nintendo Domination**:
  Nintendo dominates the sales between 2006-2010. In all regions, the total sales of Nintendo score the highest among all publishers.

- **Interpretation**:
  The data suggests that all publishers have a strong global presence in the industry, even though regional preferences for specific publishers may exist.

Insights

- This visualization highlights the dominance of Nintendo in the global gaming market. It would be interesting to explore the factors that contribute to this dominance, such as the origin of popular gaming franchises or historical trends in console manufacturing.

Top 5 Platform by Years (2006-2010)
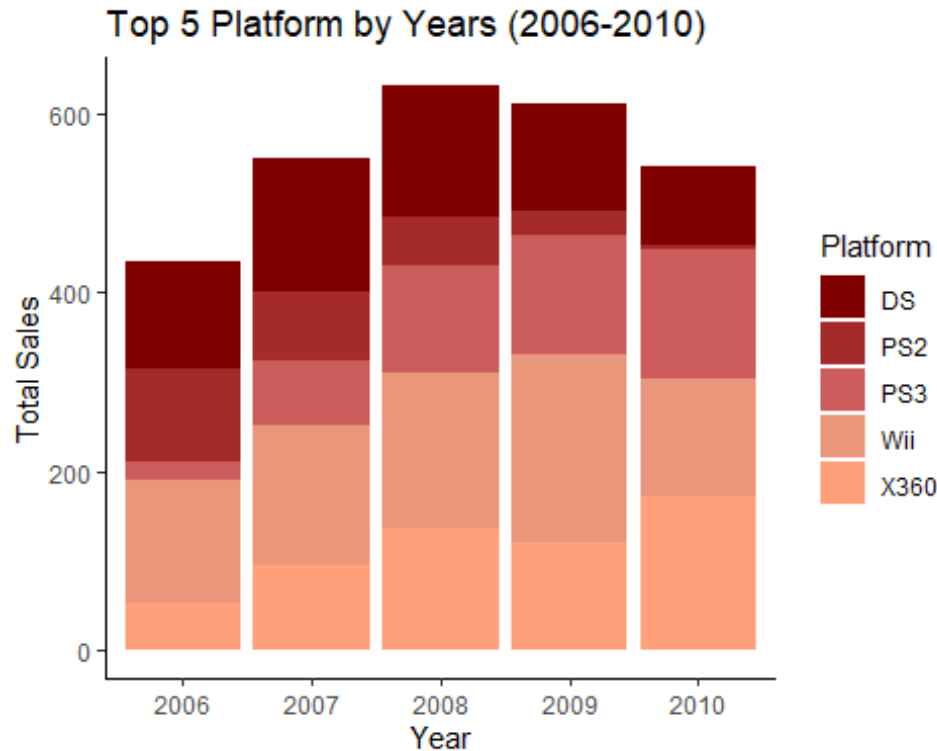
Key Observations

- **Nintendo Domination**:
  Nintendo DS and Nintendo Wii appear to be the top platform that sold the most video games throughout the five years.

- **Consistent ranking**:
  Xbox and PS3 consistently rank in the top 3 to 4 platforms throughout the five years.

- **Least popular platform**:
  PS2 appears to be the least popular platform among the five listed.

- **Interpretation**:
  The data suggests that Nintendo was the most popular video game company with their DS and Wii consoles during the 2006-2010 period.

Insights

- The visualization might indicate that during the 2006-2010 period, handheld gaming consoles were more popular than home consoles. It is also worth noting that the visualization only shows data for a specific period of time. Popularity of video game platforms can change over time.

A. Chi Square

Definition
The Chi-Square test is a statistical procedure for determining the difference between observed and expected data. This test can also be used to determine whether it correlates to the categorical variables in our data. It helps to find out whether a difference between two categorical variables is due to chance or a relationship between them [4].

A chi-square test is a statistical test that is used to compare observed and expected results. The goal of this test is to identify whether a disparity between actual and predicted data is due to chance or to a link between the variables under consideration. As a result, the chi-square test is an ideal choice for aiding in our understanding and interpretation of the connection between our two categorical variables.

Why do we use Chi Square Test?
There are a few advantages in using the Chi Square Test. Advantages of the Chi-square include its robustness with respect to distribution of the data, its ease of computation, the detailed information that can be derived from the test, its use in studies for which parametric assumptions cannot be met, and its flexibility in handling data from both two group and multiple group studies.

Chi Square Test Formula

$$\chi^2 = \Sigma \frac{(O_i - E_i)^2}{E_i}$$

Where:

- χ^2 (chi-square) = statistic test value
- i = subscript denoting individual cells in the contingency table
- O_i = observed value in cell i
- E_i = expected value in cell i

The degrees of freedom in a statistical calculation represent the number of variables that can vary in a calculation. The degrees of freedom can be calculated to ensure that chi-square tests are statistically valid. These tests are frequently used to compare observed data with data that would be expected to be obtained if a particular hypothesis were true [5].

Chi-Square test results for Year, Genre, Platform, and Publisher distributions

Test

ChiSquareStatistic

PValue

Year

98.11726

0

Genre

1745.53856

0

Platform

5331.36030

0

Publisher

377.22834

0

Chi Square Result:
The chi-square test results in the image show that there is no statistically significant correlation between the year, genre, platform, and publisher distributions of games.

- **Test** : This column shows the chi-square statistic which is a measure of how much the observed data deviates from what would be expected under the null hypothesis (i.e., there is no relationship between the variables).
  Chi Square Statistic - These values represent the chi-square statistic for each test (Year, Genre, Platform, Publisher).
- **PValue** : This column shows the p-value, which is the probability of observing a chi-square statistic this extreme or more extreme, assuming the null hypothesis is true. A low p-value (typically below 0.05) indicates that the observed data is unlikely to have occurred by chance, and thus we can reject the null hypothesis and conclude that there is a relationship between the variables.

B. Kruskal Wallis

Definition
The Kruskal–Wallis test is a statistical test used to compare two or more groups for a continuous or discrete variable. It is a non-parametric test, meaning that it assumes no particular distribution of your data and is analogous to the one-way analysis of variance (ANOVA) [6].

Why do we use Kruskal-Wallis Test?
The Kruskal Wallis test and other non-parametric (or distribution-free) tests are useful to test hypotheses when the assumption for normality of the data does not hold. They make no assumptions about the shape of data distributions, and this makes them particularly useful when a dataset is small. Our dataset is suitable for this method as our gathered data is considered relatively small.

Kruskal-Wallis Test Formula

Let's say the null hypothesis is true and thus there is no difference between the independent samples. Then high and low ranks are randomly distributed across the samples and should be equally distributed across the groups. Therefore, the probability that a rank is assigned to a group is the same for all groups [7].

If there is no difference between the groups, the mean value of the ranks should also be the same in all groups. The expected value of the ranks for each group is then given by

$$E_R = \frac{n+1}{2}$$

Each sample has the same expected value of the ranks, which corresponds to the expected value of the population. Furthermore, the variance of the ranks is needed, the variance can be calculated with the following formula:

$$\sigma^2 = \frac{n^2 - 1}{12}$$

In the Kruskal-Wallis test, the test variable H is calculated. The H value corresponds to the X^2 value. The H value results from:

$$H = \frac{n-1}{n} \sum_{i=1}^{k} \frac{(R_i - \bar{E_R})^2}{\sigma^2}$$

Kruskal-Wallis Test Results for Sales of Game Names by Region and Platform

Variable

Statistic

P_value

Global Sales

9.0000

0.437

NA Sales

9.0000

0.437

EU Sales

9.0000

0.437

JP Sales

9.0000

0.437

Other Sales

9.0000

0.437

Platform

871.4586

0.000

Kruskal Wallis Test Result:

- **Kruskal-Wallis Test Results for Sales of Game Names by Region and Platform**:

Since the p-value is more than 0.05, we can say the null hypothesis is accepted. This means that there is no statistically significant difference between the medians of sales across regions and platforms.

Kruskal-Wallis Test Results for Sales of Genres by Region

Variable

Statistic

P_value

Global Sales

9

0.437

NA Sales

9

0.437

EU Sales

9

0.437

JP Sales

9

0.437

Other Sales

9

0.437

- **Kruskal-Wallis Test Results for Sales of Genres by Region**:

Since the p-value is greater than 0.05, we accept the null hypothesis, which states that there is no statistically significant difference between the medians of sales across genres.

Kruskal-Wallis Test Results for Sales of Publishers by Region

Variable

Statistic

P_value

Global Sales

9

0.437

NA Sales

9

0.437

EU Sales

9

0.437

JP Sales

9

0.437

Other Sales

9

0.437

- **Kruskal-Wallis Test Results for Sales of Publishers by Region**:

Since the p-value is greater than 0.05, we fail to reject the null hypothesis. The null hypothesis states that there is no statistically significant difference between the medians of sales across publishers.

6. Discussion

Sales Patterns and Regional Preferences:

- Flat Sales Distribution: The median scores approaching zero across all regions indicate that most games achieve similar sales figures, suggesting market stability without extreme fluctuations for the majority of games. However, outliers indicate some games perform significantly better or worse than average.

- Regional Genre Variations: Action genre dominates consistently across all regions, but the differences are more pronounced in North America and Europe compared to Japan. This highlights the importance of considering local preferences and trends in game development and marketing.

Regional Impact on Global Sales:

- Major Market Contributions: North America contributes the most to global sales, while Japan contributes less. This underscores the importance of targeting and prioritizing major markets like North America in global marketing strategies.

- Population Factor: Larger populations in regions like North America significantly impact global sales, indicating demographic factors play a crucial role in global sales strategy.

Annual Sales Trends:

- Significant Yearly Fluctuations: There was a notable spike in sales in 2008, possibly due to external factors such as new console launches or favorable economic conditions. This pattern suggests external market conditions can have a substantial impact on sales performance over specific periods.

Platform and Genre Dominance:

- Dominant Platforms and Genres: Nintendo DS and Wii dominate the platform market, while the Action genre dominates in terms of the number of releases. This suggests focusing on dominant platforms and genres can maximize sales potential and market reach.

Player Content and Preferences:

- Popularity of Action Genre: The Action genre is globally popular, with stable demand year over year, indicating prioritizing Action game development can meet strong market demand.

- Regional Genre Variations: Preferences for specific genres, such as Role-Playing in Japan, highlight the importance of understanding and meeting cultural and regional preferences to enhance market penetration.

Publisher Dominance and Global Strategy:

- Nintendo's Dominance: Nintendo dominates global sales, demonstrating the strength of their franchises and strategies in the gaming market. This emphasizes leveraging strong brands and franchises to secure significant market share.

- Global Presence: Top publishers have strong global presence but vary in performance across regions, suggesting local adaptation and regional marketing strategies can enhance global success.

Implications for the Game Industry Strategy:

- Game Development and Marketing: Understanding regional preferences in genres and platforms can aid in designing more effective development and targeted marketing strategies.

- Adapting Market Trends: Monitoring and responding to fluctuating market trends can provide a competitive advantage, such as launching games during peak consumer interest periods.

- Prioritizing Major Markets: Focusing on major markets like North America can enhance global sales potential due to their significant contributions.

- Targeted Content: Developing content tailored to local preferences and culture can increase game acceptance and performance in diverse markets.

7. Conclusion
- The results of sales changes and popular regions in the gaming industry from 2006 to 2010 are useful in the development of a strategy. Yearly sales vary with different years having different amounts of sales influenced by new console-launches and economic conditions as depicted by the extremely high sales recorded in 2008.

- Overall, the action genre was the most popular, although the proportion of the genre in different countries varied; for example, a larger proportion of role-playing games were from Japan.

- The largest results were demonstrated by the North American region, followed by the European region, and other regions such as Japan and the rest of the world combined were less.

- Nintendo was the number one star in sales with strong franchises and the most games and top titles namely Wii Sports, Mario Kart Wii, Wii Sports Resort, and others, also in hardware as well with Nintendo DS and Wii.

- Overall, this analysis underlines the importance of understanding regional preferences, focusing on major markets, and leveraging strong brands to enhance sales in thefuture.

- Market Segmentation: Recognizing regional game preferences (e.g., Action dominance globally, Role-Playing in Japan) improves marketing effectiveness.

- Global Strategy: Prioritizing North America and Europe for launches, campaigns, and tailored content boosts global sales.
- Market Trends: Adapting to fluctuating trends (e.g., timing launches with peak consumer interest) offers a competitive edge.

- Strong Brands: Using powerful franchises (e.g., Nintendo's) to capture substantial market share.

8. Reference

[1] A. Verma, "Exploratory Data Analysis and Visualization Techniques in Data Science," Analytics Vidhya, Aug. 2021. [Online]. Available: https://www-analyticsvidhya-com.translate.goog/blog/2021/08/exploratory-data-analysis-and-visualization-techniques-in-data-science/?_x_tr_sl=en&_x_tr_tl=id&_x_tr_hl=id&_x_tr_pto=tc. [Accessed: Jun. 18, 2024].

[2] "Kruskal-Wallis H Test using SPSS Statistics," Laerd Statistics, [Online]. Available: https://statistics-laerd-com.translate.goog/spss-tutorials/kruskal-wallis-h-test-using-spss-statistics.php?_x_tr_sl=en&_x_tr_tl=id&_x_tr_hl=id&_x_tr_pto=tc. [Accessed: Jun. 18, 2024].

[3] "Chi-Square Test," Simplilearn, [Online]. Available: https://www-simplilearn-com.translate.goog/tutorials/statistics-tutorial/chi-square-test?_x_tr_sl=en&_x_tr_tl=id&_x_tr_hl=id&_x_tr_pto=tc. [Accessed: Jun. 18, 2024].

[4] "Chi-Square Test," Simplilearn, [Online]. Available: https://www.simplilearn.com/tutorials/statistics-tutorial/chi-square-test. [Accessed: Jun. 18, 2024].

[5] A. M. Luna et al., "Advantages of the Chi-Square Test," PubMed, Jul. 2013. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/23894860/#:~=Advantages%20of%20the%20Chi%2Dsquare,both%20two%20group%20and%20multiple. [Accessed: Jun. 18, 2024].

[6] "The Kruskal-Wallis Test," Technology Networks, [Online]. Available: https://www.technologynetworks.com/informatics/articles/the-kruskal-wallis-test-370025. [Accessed: Jun. 18, 2024].

[7] "Kruskal-Wallis Test," DataTab, [Online]. Available: https://datatab.net/tutorial/kruskal-wallis-test. [Accessed: Jun. 18, 2024].