1 Introduction

The purpose of this research is to examine trends in the video game business by looking at the relationships between game genres, critical and user ratings, and sales numbers. Furthermore, the study dives into the market's preferences for gameplay-centric versus story-centric video games, as well as how these preferences have evolved over time.

The study makes use of a dataset compiled from several internet sites that includes video games released between 1980 and 2016. Details such as video game titles, genres, critical and user ratings, sales numbers, and release dates are included in this dataset.

The goal of this data research is to provide significant insights about the video game sector that can assist game developers, investors, and players alike. Furthermore, the findings may reveal possible chances for future video game creation and aid in understanding the shifting preferences of video game enthusiasts.

2 Data used

2.1 Data source

The work presented here is based on the Tagged_Data_Final.csv file from Global Video Game Sales & Ratings Dataset, hosted on Kaggle.

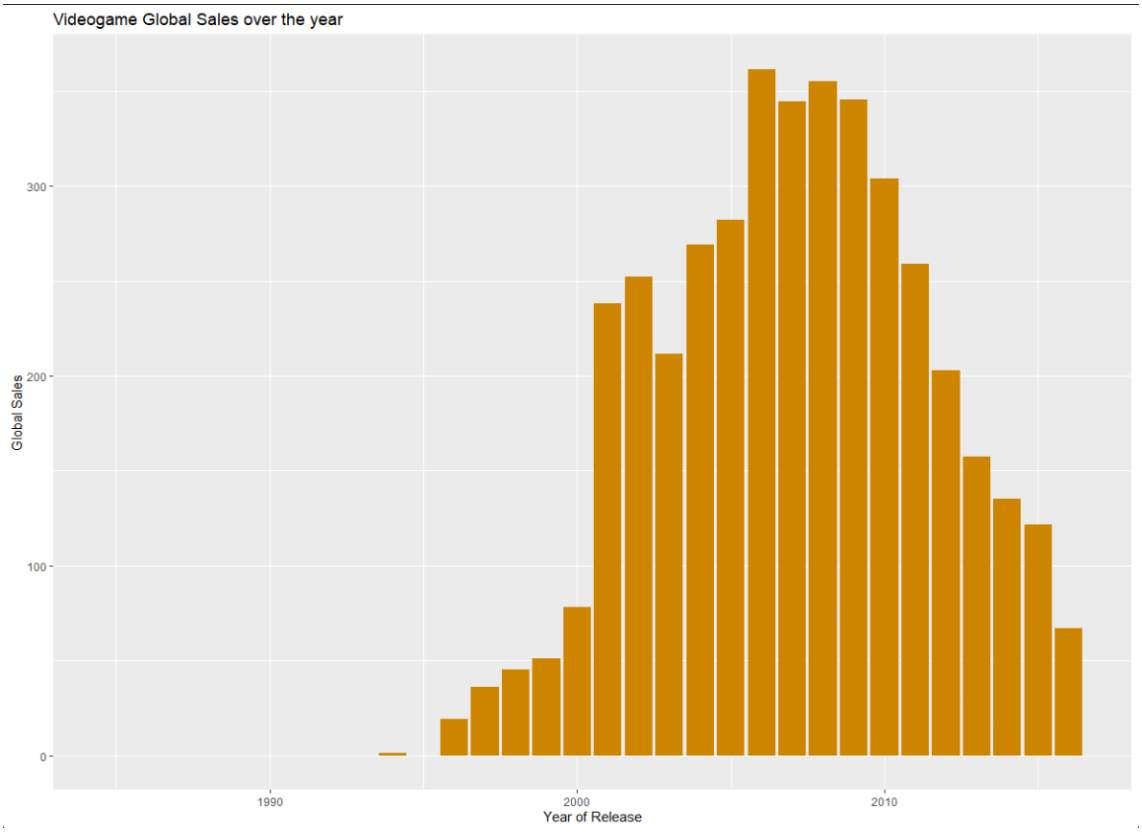
2.2 Dataset structure

The dataset is organized with the following structure: it includes columns for "Name," "Year_of_Release," "Genre," "Publisher," "NA_Sales," "EU_Sales," "JP_Sales," "Other_Sales," and "Global_Sales," providing information on regional and global sales figures. Additionally, it incorporates "Critic_Score," "Critic_Count," "User_Score," and "User_Count" to capture both critical and user assessments. The dataset also encompasses details on the game's "Developer," "Rating," and specifies whether the game focuses on "Story" or "Gameplay." Lastly, there is a column labelled "Series" to indicate if the game belongs to a series. This comprehensive structure ensures a rich and varied dataset, offering insights into various aspects of video games, from sales performance to critical and user reception, as well as other relevant attributes.

3 Videogames market

3.1 Global sales

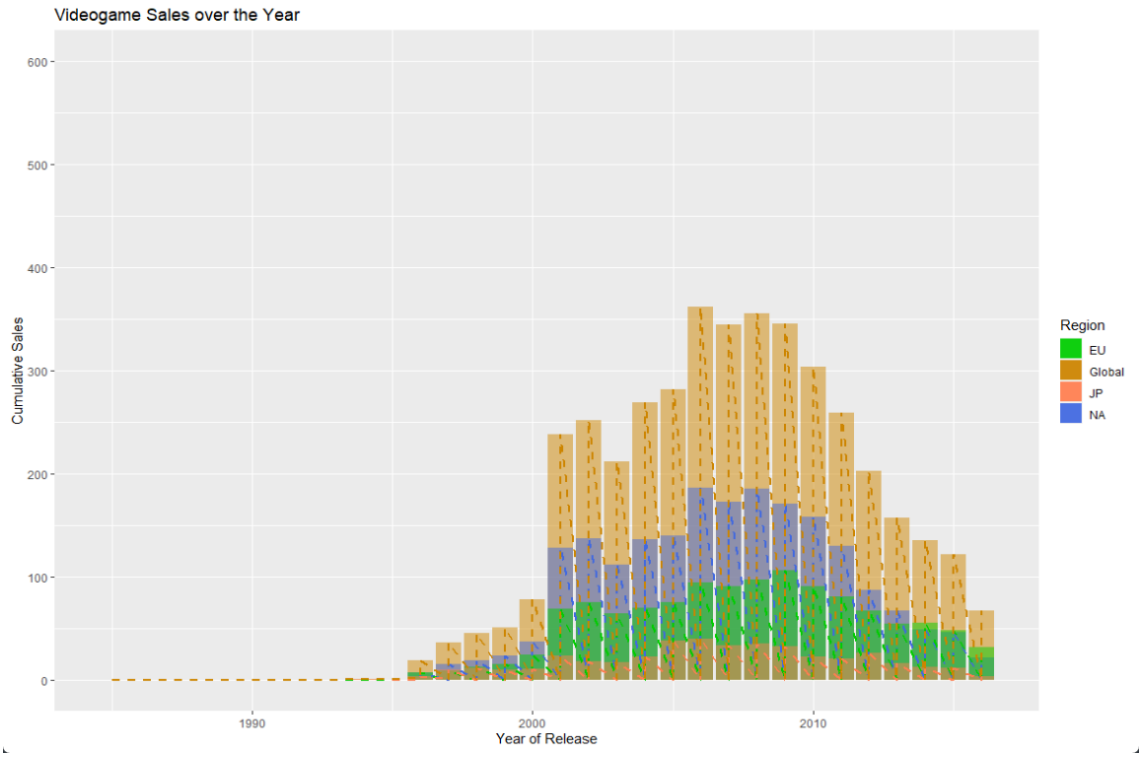
```
# Videogame sales trends per year
ggplot(final_dataset, aes(year)) +
  geom_col(aes(y = prev_global_sales), fill = "orange3") +
  labs(title = "Videogame Global Sales over the year",
       x = "Year of Release", y = "Global Sales")

```



As shown in the graph global sales reached their peak after 2005 and after 2010 the decline of sales started to decline.

3.2 Videogame sales over the years in key regions

```
#comparison between global sales and regional sales of North America, Europe and Japan
ggplot(final_dataset, aes(x = year)) +
  geom_col(aes(y = prev_global_sales, fill = "Global"), alpha = 0.5) +
  geom_col(aes(y = prev_NA_sales, fill = "NA"), alpha = 0.5) +
  geom_col(aes(y = prev_EU_sales, fill = "EU"), alpha = 0.5) +
  geom_col(aes(y = prev_JP_sales, fill = "JP"), alpha = 0.5) +
  stat_summary(aes(y = prev_NA_sales), fun = cumsum, geom = "line", linetype = "dashed", color = "royalblue", size = 1) +
  stat_summary(aes(y = prev_EU_sales), fun = cumsum, geom = "line", linetype = "dashed", color = "green3", size = 1) +
  stat_summary(aes(y = prev_JP_sales), fun = cumsum, geom = "line", linetype = "dashed", color = "coral", size = 1) +
  stat_summary(aes(y = prev_global_sales), fun = cumsum, geom = "line", linetype = "dashed", color = "orange3", size = 1) +
  labs(title = "Videogame Sales over the Year",
       x = "Year of Release",
       y = "Cumulative Sales",
       fill = "Region") +
  scale_fill_manual(values = c("Global" = "orange3", "NA" = "royalblue", "EU" = "green3", "JP" = "coral")) +
  scale_y_continuous(limits = c(0, 600), breaks = seq(0, 600, by = 100))

```

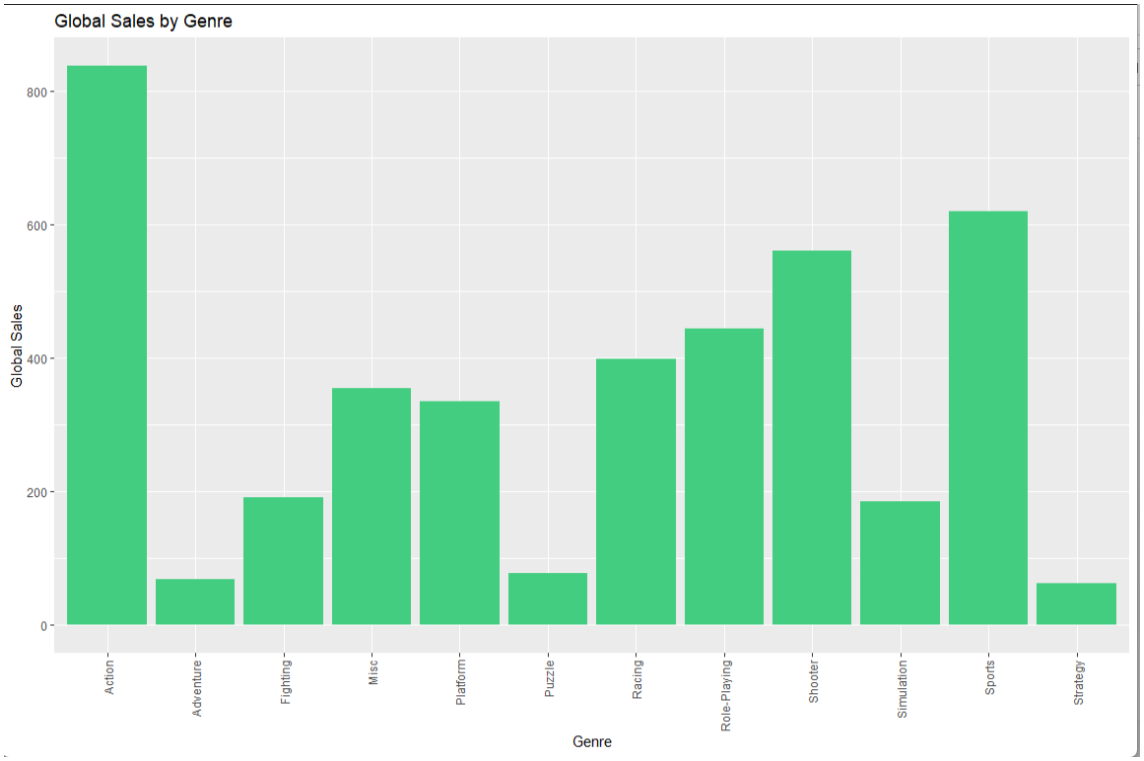


While Japan historically had the lowest sales volume, Europe's sales surpassed it by more than double. In contrast, North American sales were twice those of Europe until 2012, after which they achieved a more balanced distribution.

It's noteworthy that Japan has consistently represented the smallest market in terms of sales. Since 2012, however, European sales have been comparable to those in North America.

3.3 Global Sales by Genre

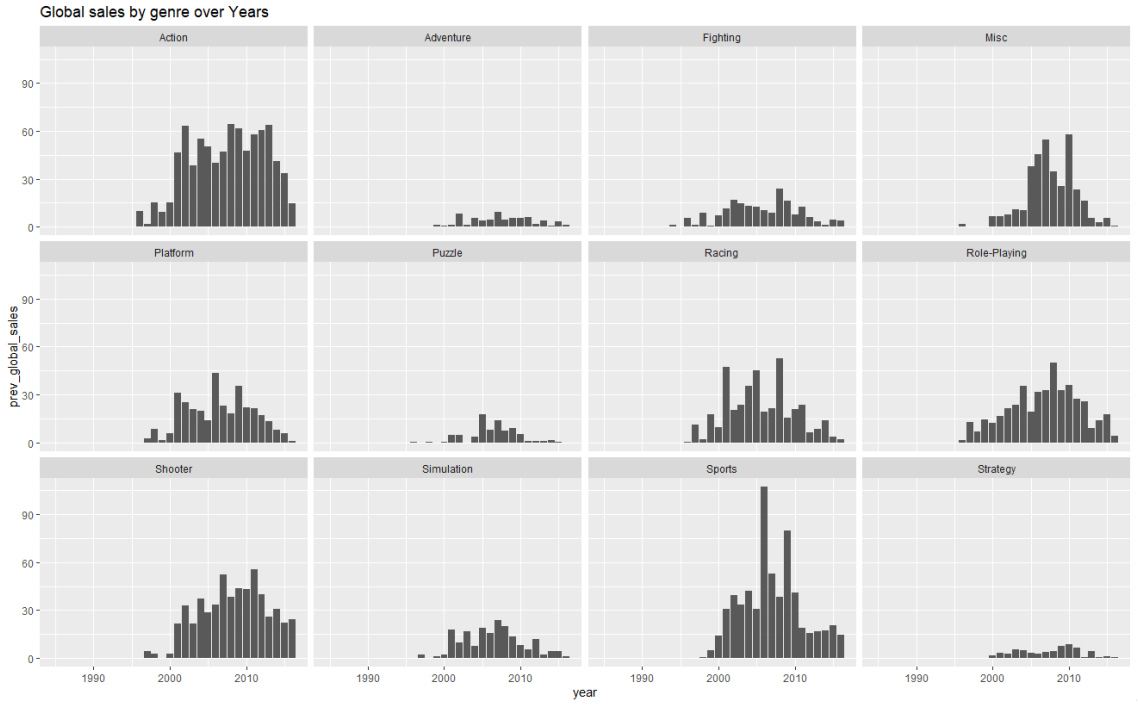
```
100  # Analysis by genre
101  ggplot(final_dataset, aes(x = genre, y = prev_global_sales)) +
102    geom_col(fill = "seagreen3") +
103    labs(title = "Global Sales by Genre",
104         x = "Genre",
105         y = "Global Sales") +
106    theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1))
107
```



As shown in the graph above the best-selling category is Action followed by Sport and Shooter. However, the categories that perform the worst in sales are Strategy, Adventure and Puzzle.

3.4 Sales by Genre over the years Global

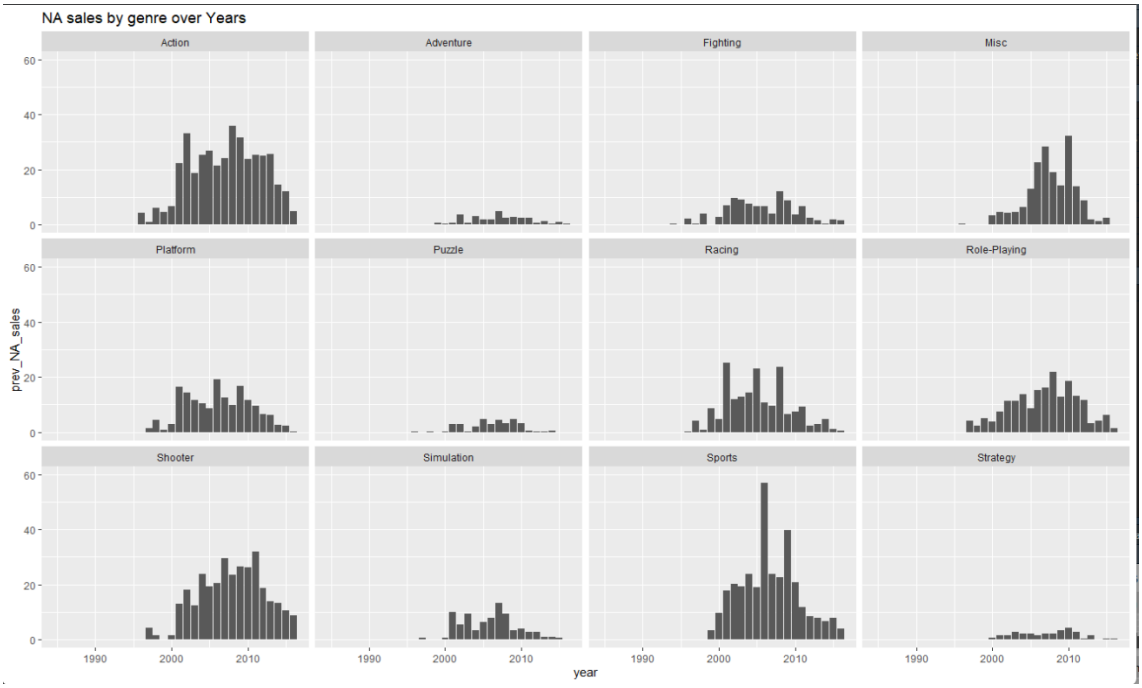
```
# Global Sales by genre over years
ggplot(final_dataset, aes(x = year, y = prev_global_sales)) +
  geom_bar(stat = "identity", position = position_stack()) +
  facet_wrap(~genre, ncol = 4) +
  labs(title = "Global sales by genre over Years")

```



The genre with the highest concentration of sales density was Action. Furthermore, large sales increases in the Sport genre were noted between 2005 and 2010.

3.5 Sales by Genre over the years in NA

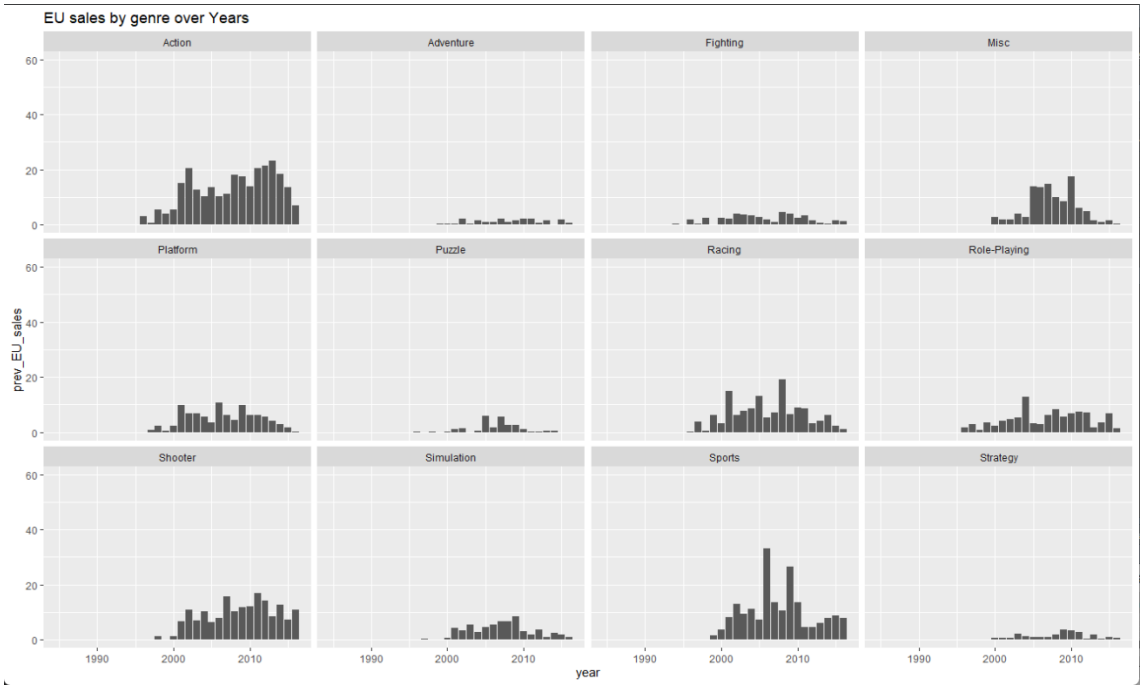
```
116  # NA sales by genre over years
117  ggplot(final_dataset, aes(x = year, y = prev_NA_sales)) +
118    geom_bar(stat = "identity") +
119    facet_wrap(~genre, ncol=4) +
120    labs(title="NA sales by genre over Years") +
121    ylim(0, 60)
122
```



Examining regional sales data for North America reveals a trend like the worldwide market. The genres with the highest sales figures are Action, Mics, and Shooter. The genres with the most concentrated sales structures are Action, Platform, Role-Playing, Sports, and Shooter. Furthermore, between 2005 and 2010, the Sport genre saw significant sales increases.

3.6 Sales by Genre over the years in EU

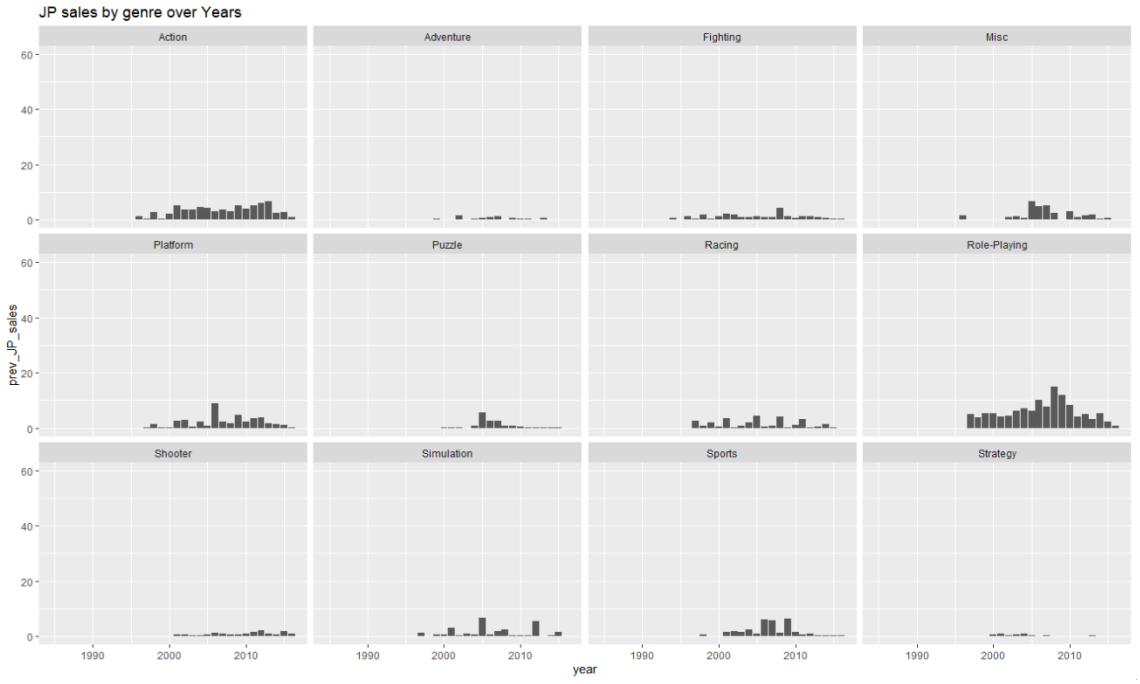
```
124  # EU sales by genre over years
125  ggplot(final_dataset, aes(x = year, y = prev_EU_sales)) +
126    geom_bar(stat = "identity") +
127    facet_wrap(~genre, ncol=4) +
128    labs(title="EU sales by genre over Years") +
129    ylim(0, 60)
130
```



We see a similar tendency when we look at data from the EU market. The Action, Platform, Miscellaneous, and Shooter genres have the highest sales density, whereas the Sport genre has noticeable jumps between 2005 and 2010. These findings are consistent with the substantial positive connection in sales numbers observed in the North American NA, EU, and global markets. This implies that the EU market follows a pattern like the NA and worldwide markets.

3.7 Sales by Genre over the years in JP

```
132  # JP sales by genre over years
133  ggplot(final_dataset, aes(x = year, y = prev_JP_sales)) +
134    geom_bar(stat = "identity") +
135    facet_wrap(~genre, ncol=4) +
136    labs(title="JP sales by genre over Years") +
137    ylim(0, 60)
138
```



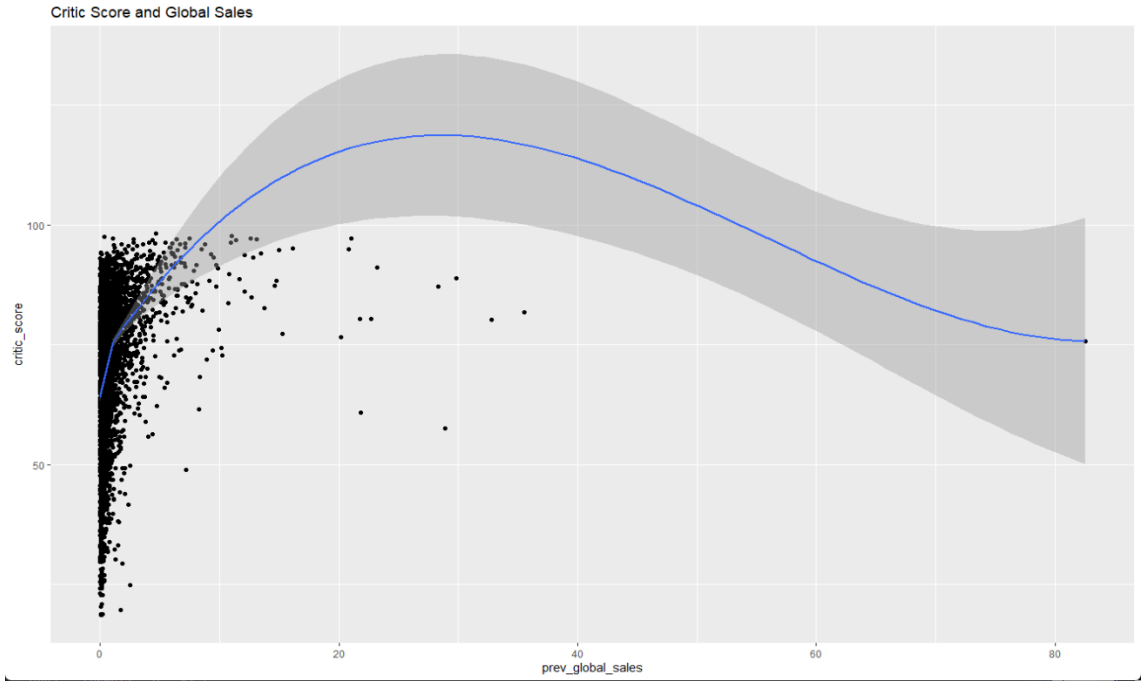
The situation in the Japanese market is very peculiar, where maximum average sales volumes for genres are smaller than in other regions. Role-Playing, Action, and Platform are the top three genres in terms of sales volume. Notably, all genres experienced significant sales increases between 2000 and 2010.

These findings imply that market preferences in Japan differ from those in other parts of the world. This distinction is further confirmed by the lower level of positive connection between Japanese sales and worldwide sales as compared to the correlation between sales in the United States and Europe and global sales.

4 Correlation between User score, critic score and sales

4.1 Critic Scores and Global Sales relation

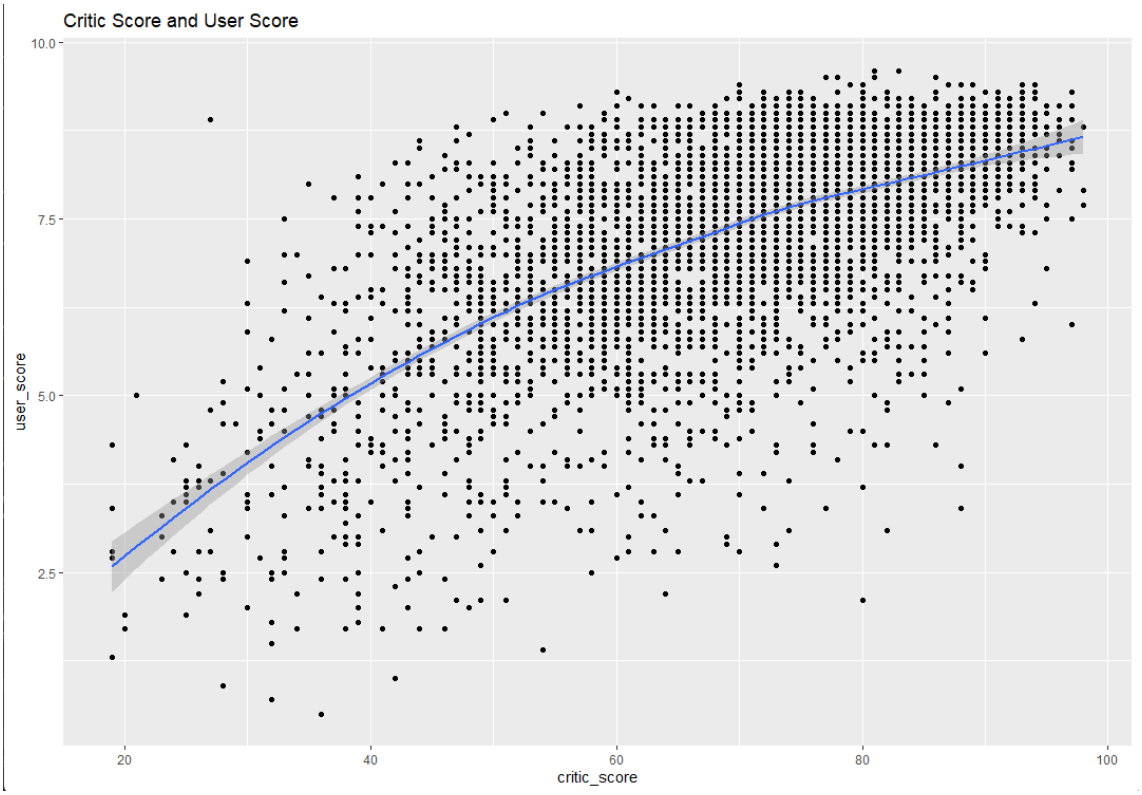
```
# Critic Scores and Global Sales relation
ggplot(final_dataset, aes(x = prev_global_sales, y = critic_score), shape = carrier) +
  geom_jitter() +
  geom_smooth(method = "loess", formula = y ~ x) +
  labs(title="Critic Score and Global Sales")

```



Examining the scatterplot reveals a positive association between critical ratings and sales, meaning that games with higher ratings do better in terms of sales. Notably, after critical evaluations exceed 60, this correlation becomes significantly stronger. The plot does, however, reveal a significant number of games with lower critical ratings, ranging from 60 to 90, but reach impressive sales statistics. These findings indicate the presence of influential outliers in the dataset, implying that factors other than critical reviews may influence game sales.

4.2 Critic Scores and User Score relation

```
# Critic Scores and User Score relation
ggplot(final_dataset, aes(x = critic_score, y = user_score), shape = carrier) +
  geom_point() +
  geom_smooth(method = "loess", formula = y ~ x) +
  labs(title="Critic Score and User Score")

```

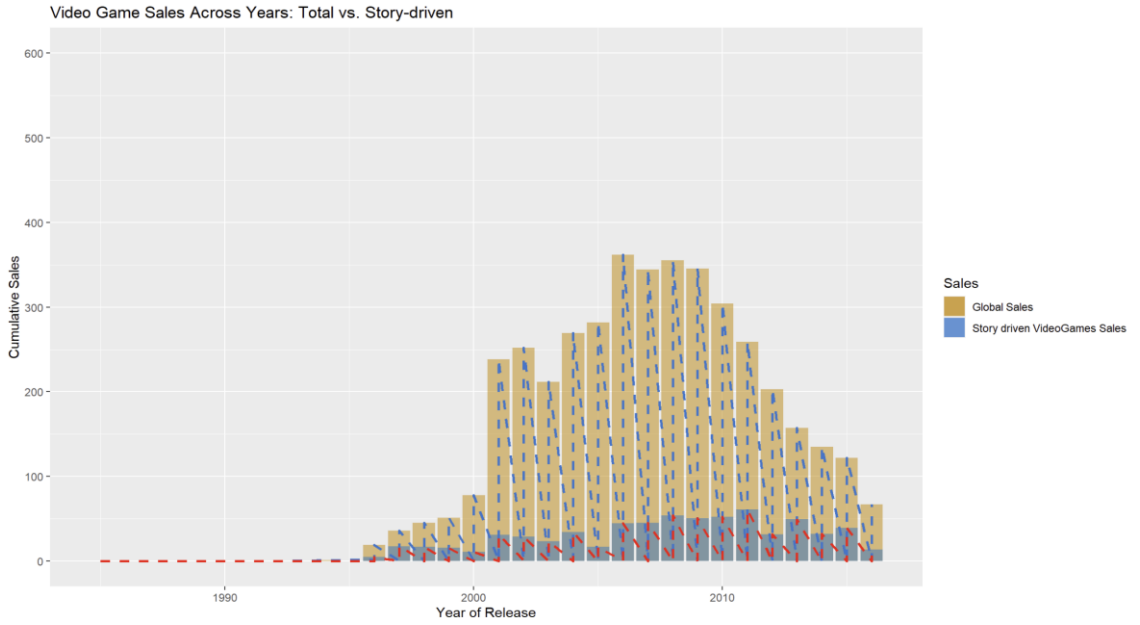


It is possible to conclude that there is a positive relationship between rating values and unanimity among reviewers and users. This association becomes stronger as the rating values increase, peaking at higher rating levels. This pattern may be seen starting with a critic rating of 60 and a user rating of 6. However, despite the general improved trend, the dispersion of data points remains significant, showing a significant range in viewpoints.

5 The trend in the market for story driven videogame

5.1 Correlation between: Story driven games and global sales

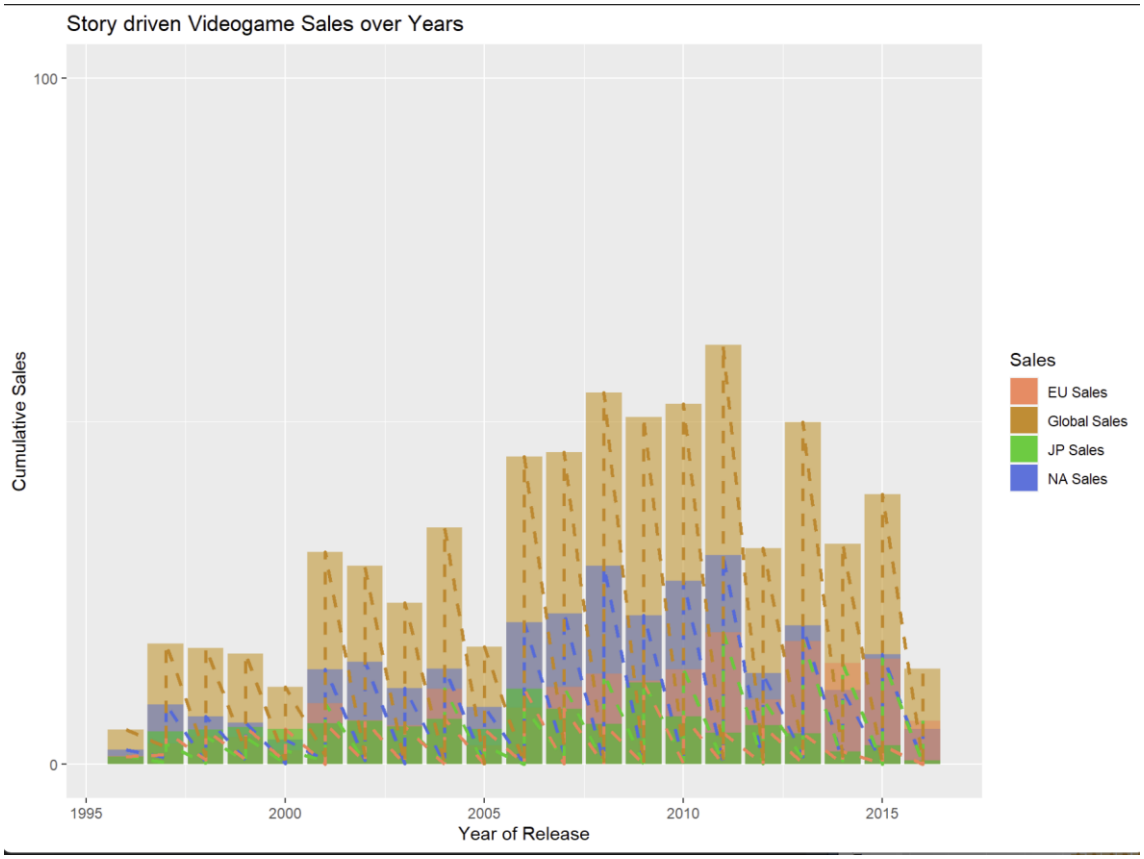
```r
# Global sales vs global story_game sales
ggplot(data = final_dataset, aes(x = year)) +
  geom_col(aes(y = prev_global_sales, fill = "Global Sales"), alpha = 0.5) +
  geom_col(aes(y = ifelse(story_focus == 1, prev_global_sales, 0), fill = "Story driven VideoGames Sales"), alpha = 0.5) +
  stat_summary(aes(y = prev_global_sales), fun = cumsum, geom = "line", linetype = "dashed", color = "dodgerblue3", linewidth = 1) +
  stat_summary(aes(y = ifelse(story_focus == 1, prev_global_sales, 0)), fun = cumsum, geom = "line", linetype = "dashed", color = "red", linewidth = 1) +
  labs(title = "Video Game Sales Across Years: Total vs. Story-driven",
       x = "Year of Release",
       y = "Cumulative Sales",
       fill = "Sales") +
  scale_fill_manual(values = c("Global Sales" = "orange3", "Story driven VideoGames Sales" = "dodgerblue3")) +
  scale_y_continuous(limits = c(0, 600), breaks = seq(0, 600, by = 100))

```



The plot reveals a clear and robust positive correlation between story-driven games and global sales, comprising around 10% of the market. Unlike global sales, story-focused video games demonstrate lower variability, displaying a tendency to be less susceptible to abrupt spikes or declines.

5.2 Trends in Story driven videogames

```r
#Story driven Videogame Sales over Years
ggplot(story_driven, aes(x = year)) +
  geom_col(aes(y = prev_global_sales, fill = "Global Sales"), alpha = 0.5) +
  geom_col(aes(y = prev_NA_sales, fill = "NA Sales"), alpha = 0.5) +
  geom_col(aes(y = prev_EU_sales, fill = "EU Sales"), alpha = 0.5) +
  geom_col(aes(y = prev_JP_sales, fill = "JP Sales"), alpha = 0.5) +
  stat_summary(aes(y = prev_NA_sales), fun = cumsum, geom = "line", linetype = "dashed", color = "royalblue", size = 1) +
  stat_summary(aes(y = prev_EU_sales), fun = cumsum, geom = "line", linetype = "dashed", color = "green3", size = 1) +
  stat_summary(aes(y = prev_JP_sales), fun = cumsum, geom = "line", linetype = "dashed", color = "coral", size = 1) +
  stat_summary(aes(y = prev_global_sales), fun = cumsum, geom = "line", linetype = "dashed", color = "orange3", size = 1) +
  labs(title = "Story driven Videogame Sales over Years",
       x = "Year of Release",
       y = "Cumulative Sales",
       fill = "Sales") +
  scale_fill_manual(values = c("Global Sales" = "orange3", "NA Sales" = "royalblue", "EU Sales" = "coral", "JP Sales" = "green3")) +
  scale_y_continuous(limits = c(0, 100), breaks = seq(0, 100, by = 100))

```

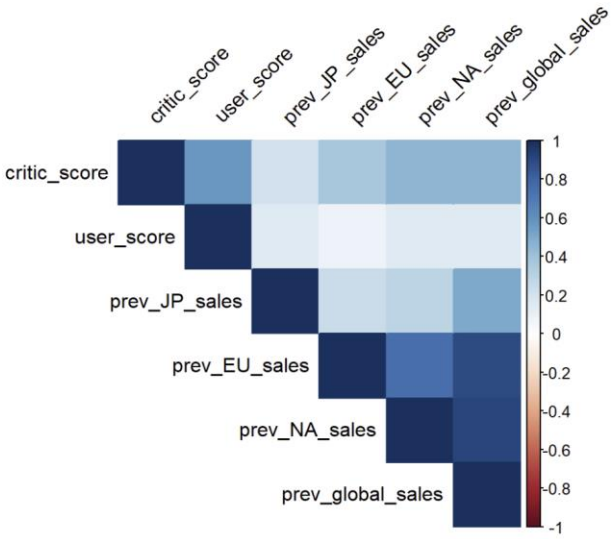


The current graph illustrates the sales of story-focused video games globally and across various regions. It mirrors overarching sales trends, including the correlation between global and regional sales, lower sales in Japan, and a shift in sales dominance from North America to Europe. However, it also reveals distinct characteristics:

- A broader and more dynamic distribution.
- Sustained sales in Japan from 1997 to 2005, followed by fluctuations from 2005 to 2010 and a subsequent decline.
- Conversely, the European market exhibits a consistent upward trend since 2010, eventually reaching parity with North America in terms of sales.

5.3 Correlation between all variables

```
# Correlation matrix
correlation_matrix <- cor(story_driven[c("prev_NA_sales", "prev_EU_sales", "prev_JP_sales", "prev_global_sales", "critic_score", "user_score")])
print(correlation_matrix)
corrplot(correlation_matrix, method="color", type="upper", order="hclust", tl.col="black", tl.srt=45)

```



The analysis indicates a robust positive correlation between story-driven games and global sales (0.90), as well as sales in North America (0.90) and Europe (0.85). However, the correlation in Japan is moderately strong (0.62), implying distinct player preferences in this region compared to other parts of the world.

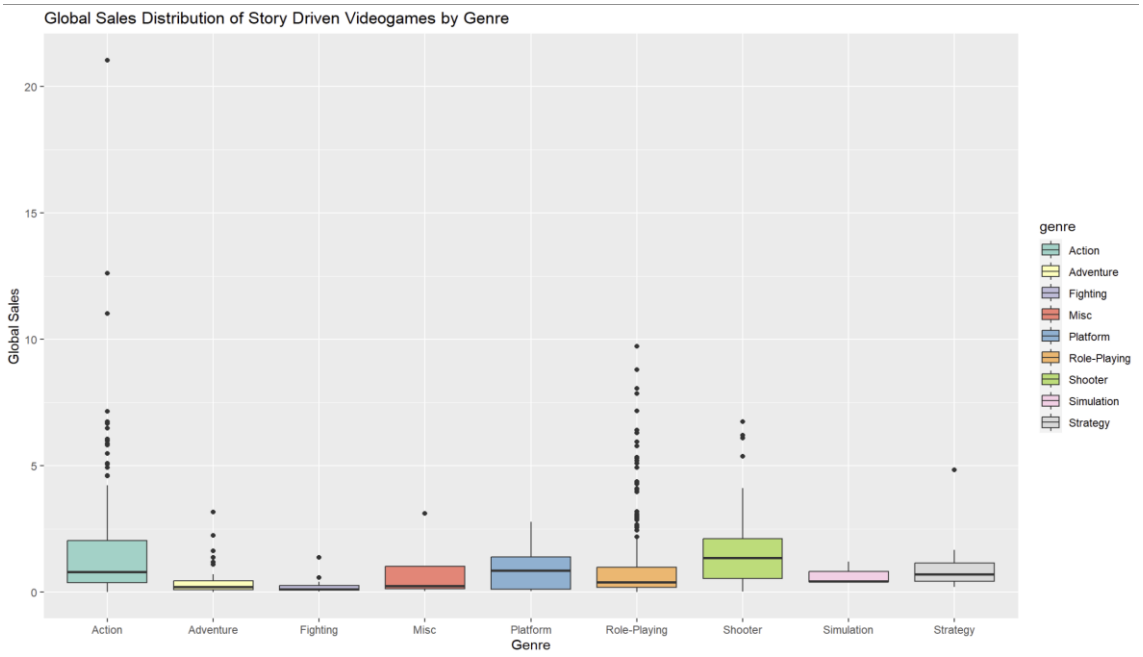
Moreover, the correlation between critic ratings and sales for historically oriented games is moderate in North America (0.46), Europe (0.38), and globally (0.45), but notably weaker in Japan (0.211). Conversely, the relationship between user ratings and sales is low, with correlation coefficients of 0.18, 0.10, 0.16, and 0.19 in each respective region.

Of note, the correlation matrix highlights a moderate positive correlation of 0.57 between critic and user ratings. This suggests that critics and users share moderately similar perspectives on the quality of historically oriented games.

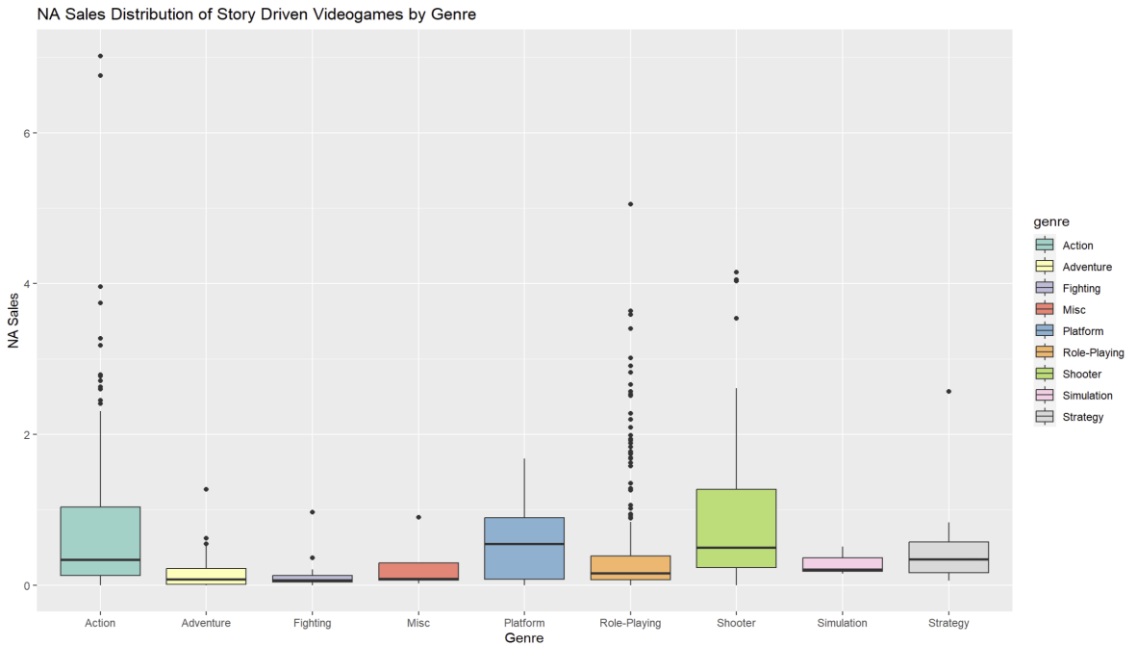
In summary, the findings affirm a strong connection between sales of story-driven games and global sales, especially in North America and Europe, with a weaker association in Japan. Additionally, critic ratings exhibit a moderate relationship with sales, while user ratings show little to no correlation with sales.

6 Genre-Specific insights of Story driven videogames

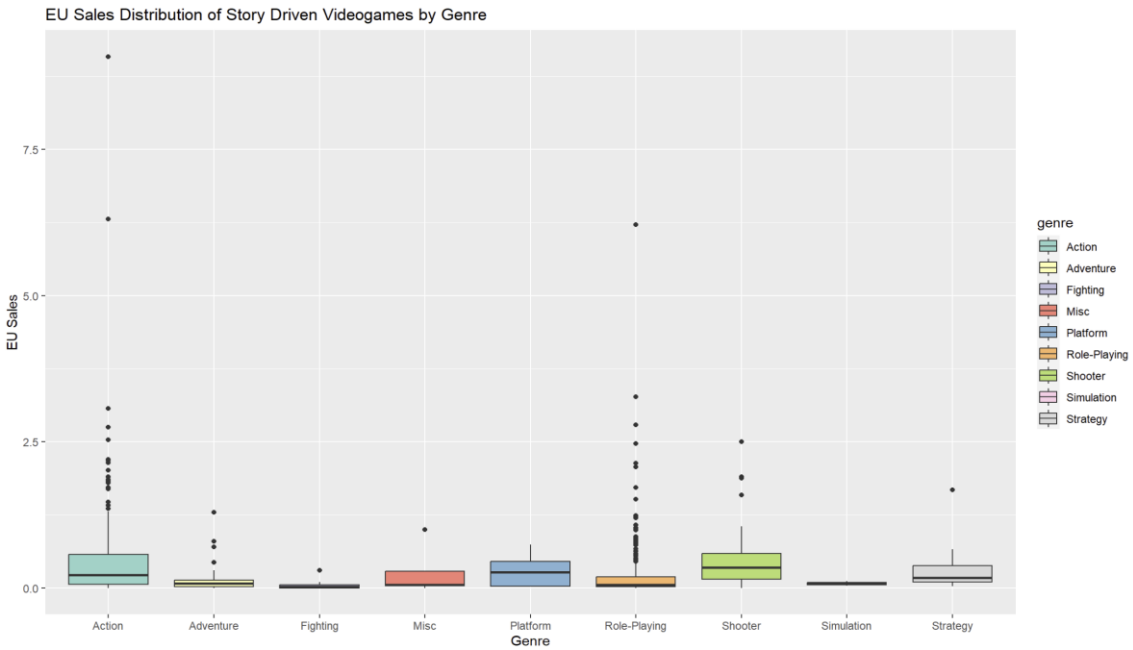
```
# Boxplot of Sales Distribution by genre
ggplot(story_driven, aes(x=genre, y=prev_global_sales, fill=genre)) +
  geom_boxplot() +
  scale_fill_brewer(palette = "Set3") +
  labs(title = "Global Sales Distribution of Story Driven Videogames by Genre",
       x = "Genre",
       y = "Global Sales")

```



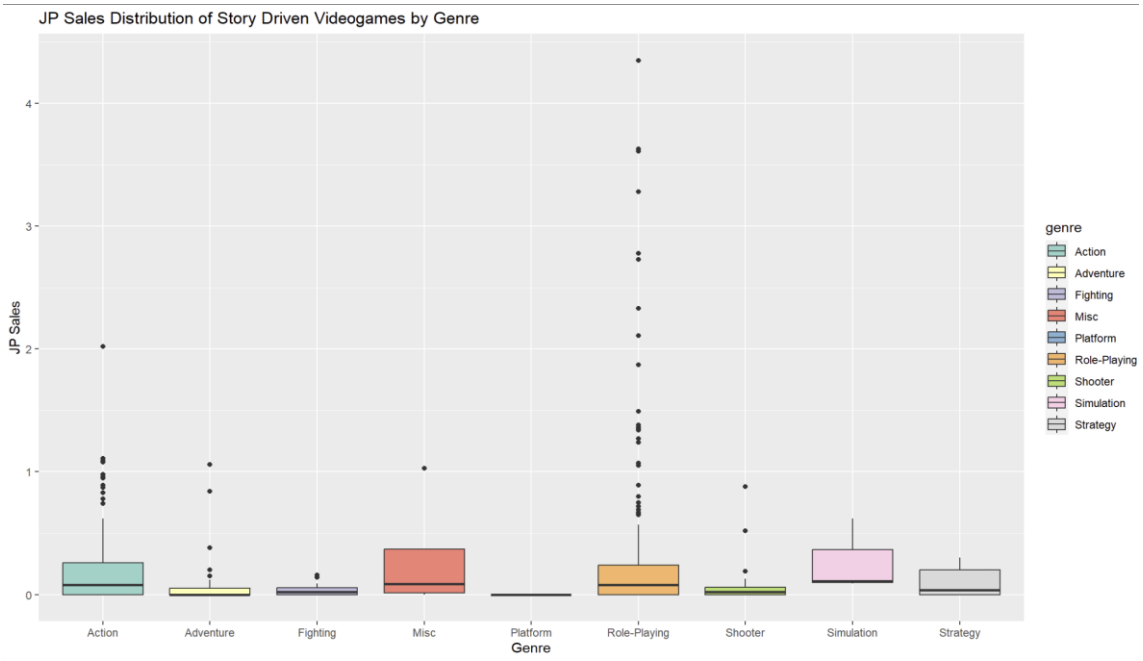

```
ggplot(story_driven, aes(x=genre, y=prev_NA_sales, fill=genre)) +
  geom_boxplot() +
  scale_fill_brewer(palette = "Set3") +
  labs(title = "NA Sales Distribution of Story Driven Videogames by Genre",
       x = "Genre",
       y = "NA Sales")

```




```
ggplot(story_driven, aes(x=genre, y=prev_EU_sales, fill=genre)) +
  geom_boxplot() +
  scale_fill_brewer(palette = "Set3") +
  labs(title = "EU Sales Distribution of Story Driven Videogames by Genre",
       x = "Genre",
       y = "EU Sales")

```




```
ggplot(story_driven, aes(x=genre, y=prev_JP_sales, fill=genre)) +
  geom_boxplot() +
  scale_fill_brewer(palette = "Set3") +
  labs(title = "JP Sales Distribution of Story Driven Videogames by Genre",
       x = "Genre",
       y = "JP Sales")

```



The box and whisker plot visually represents the distribution of video game sales across different genres, using key statistical measures. Key observations from the analysis include:

1. Action Genre: It exhibits the highest median sales among all genres with a wide range. However, its decreasing lower sales limit indicates heightened competition from adjacent genres. The scattered dots suggest high sales variability.
2. Platform Genre: While having a large box, it lacks vertical dots beyond the median, implying less diverse sales.
3. Role-Playing Genre: It displays numerous outliers for high sales, indicating a broad audience. Although having a high median sale, the sales values at the top are relatively smaller, suggesting more evenly distributed sales.
4. Shooter Genre: It has fewer outliers for high sales, suggesting a narrower audience, and a smaller box compared to the Action genre.
5. Mix Genre: This genre has the fewest sales and a low median, with only one dot on the chart, indicating limited sales diversity.

The analysis of sales dynamics is similarly observed in both North America and Europe. The Action genre dominates with high median sales but faces increased competition,

while the Shooter genre indicates a narrower audience. Role-Playing genres show high median sales with more evenly distributed sales.

In Japan, the Misc genre stands out as the top-selling genre with high sales but limited diversity. The Action genre also plays a significant role, showing increased sales variability. The Role-Playing genre is crucial, exhibiting both high sales and variability. Other genres have lower sales and less variability.

In summary, the box and whisker plot provide valuable insights into sales dynamics across different genres in North America, Europe, and Japan. It highlights genre-specific trends, competition levels, and audience diversity.

7 Conclusion

The global video game market has experienced a decline in sales since 2012, reaching its peak between 2000 and 2010. Notably, the Action, Sports, and Shooter genres consistently lead in sales, while Puzzle, Adventure, and Strategy genres demonstrate lower sales volumes. However, in the Japanese market, the top-selling genre is Role-Playing. Sales in North America and Europe have been comparable since 2012, with the highest sales attributed to the Action, Platform, Misc, and Shooter genres.

Games with higher ratings, both from critics and users, tend to perform better, indicating a positive correlation between critical ratings and sales and a weak to moderate positive correlation between user ratings and sales. However, the presence of influential outliers and additional factors beyond ratings suggests a nuanced impact on game sales. While ratings serve as useful indicators of a game's potential success, other elements also play a role.

In the overall market, gameplay-focused games dominate, with a substantial interest in story-focused games, particularly in the Role-Playing and Action genres. The prime period for story-focused games was between 2005 and 2012, followed by a gradual decline in releases. Despite the prevalence of gameplay-focused titles, a significant number of games strike a balance between gameplay and narrative.

Examining story-driven games, a notable positive correlation with global sales is observed, constituting around 10% of the market. Role-Playing games lead in popularity, closely followed by Action and Shooter genres. While Role-Playing games are favored universally, regional preferences vary for Action and Shooter games, with Europe showing a higher demand for Action, North America favoring Shooter games, and Japan dominated by the Role-Playing genre.

The box and whisker diagram highlight that the Action genre boasts the highest median sales, trailed by Role-Playing and Platform games. However, the decreasing lower sales limit in the Action genre suggests rising competition in adjacent genres. Overall, video game sales exhibit strong connections with global sales, particularly in North America and Europe, while Japan reflects distinct player preferences.

Furthermore, critic ratings moderately influence game sales, while user ratings exhibit little to no correlation with sales.