

Video Game Analysis - Final Project

Daniel Borrelli

04-19-2023

Attention Project Manager: Mr. Stone

I am responding to your request for analysis of video games statistics spanning 1980 to 2016. I believe that this report will enable our division to potentially identify where we may want to allocate our resources for the launch of our upcoming new video game releases.

Description:

After conducting a search for a data set, I located a comma separated variable (.csv) file that provided a comprehensive list of video games statistics spanning from 1980 to 2016. After reviewing and verifying the data, I felt that this data set contained significant information regarding the genre, gaming platforms, publisher, and sales data across geographic regions of North America, Europe, Japan, and the globe to serve for project presentation. Finally, Mr. Stone because we have previously discussed our love for video games and both being child of the 1980's when the personal gaming platform industry became an important part of after school entertainment I enjoyed with friends and siblings we now have data to present to that will enable us to fill the need of new gamers with our upcoming releases.

Processing problems:

In processing the data set for video game sales, I first used Excel software to create pivot table that would enable me to quickly interpret the data. After an examination of the data in an Excel Pivot table, I quickly reviewed the data across multiple geographic regions, which I found to be interesting for my final project. When the final project was assigned, I felt my prior experience with the Excel pivot table tool enabled me to overcome my inexperience with R Studio for understanding the data set. However, using Excel I quickly determined that the genre of video games in the North American market that sales are the highest are sports games, which are the highest selling games in the North American market, which account for approximately 19% of total sales between 2006 and 2009.

Big Question:

I intend to use data analytics software such as R Studio to conduct research on video game statistics spanning 1980 to 2016. I will attempt to uncover and interpret the following key questions in my analysis. First, in what years did the sales of video games reach their peak in the US market? Second, what genre of video games are the highest in North American market with sales spanning a 2009-to-2016-year period? Finally, what video game genre have the highest sales in Japan, North America, Europe and in the Global market? All these questions will be used to summarize the dataset for video game sales across multiple years and provide value to business for future video game releases. Additionally, this dataset can be used to ensure that ethical dilemmas are being considered when appropriately choosing whether we should release additional shooter style video games, a concern that our executive management and shareholders have started to express their concerns recently. However, when considering this concern, I was able to analyze additional markets and

genres that we could direct our video game sales towards.

Results:

```
library(ggplot2)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##   filter, lag
```

```
## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
```

```
library(tidyr)

VGS <- read.csv("vgsales.csv")

VGS2 <- read.csv("vgsales.csv")

VGS3 <- read.csv("vgsales.csv")

VGS2 <- VGS2[!(VGS2$Year %in% c("N/A", "2017", "2020")),]

VGS3 <- VGS3[!(VGS3$Year %in% c("N/A", "2017", "2020")),]

VGS <- VGS[!(VGS$Year %in% c("N/A", "2017", "2020")),]
VGS <- VGS %>% gather(Region, Revenue, 7:10)
VGS$Region <- factor(VGS$Region)

# Function to define the theme use across all the plots in the data analysis
theme_1 <- function() {

  return(theme(axis.text.x = element_text(angle = 90, size = 12, vjust = 0.4), plot.title = el
ement_text(size = 16, vjust = 2),axis.title.x = element_text(size = 12, vjust = -0.33)))

}

theme_2 <- function() {

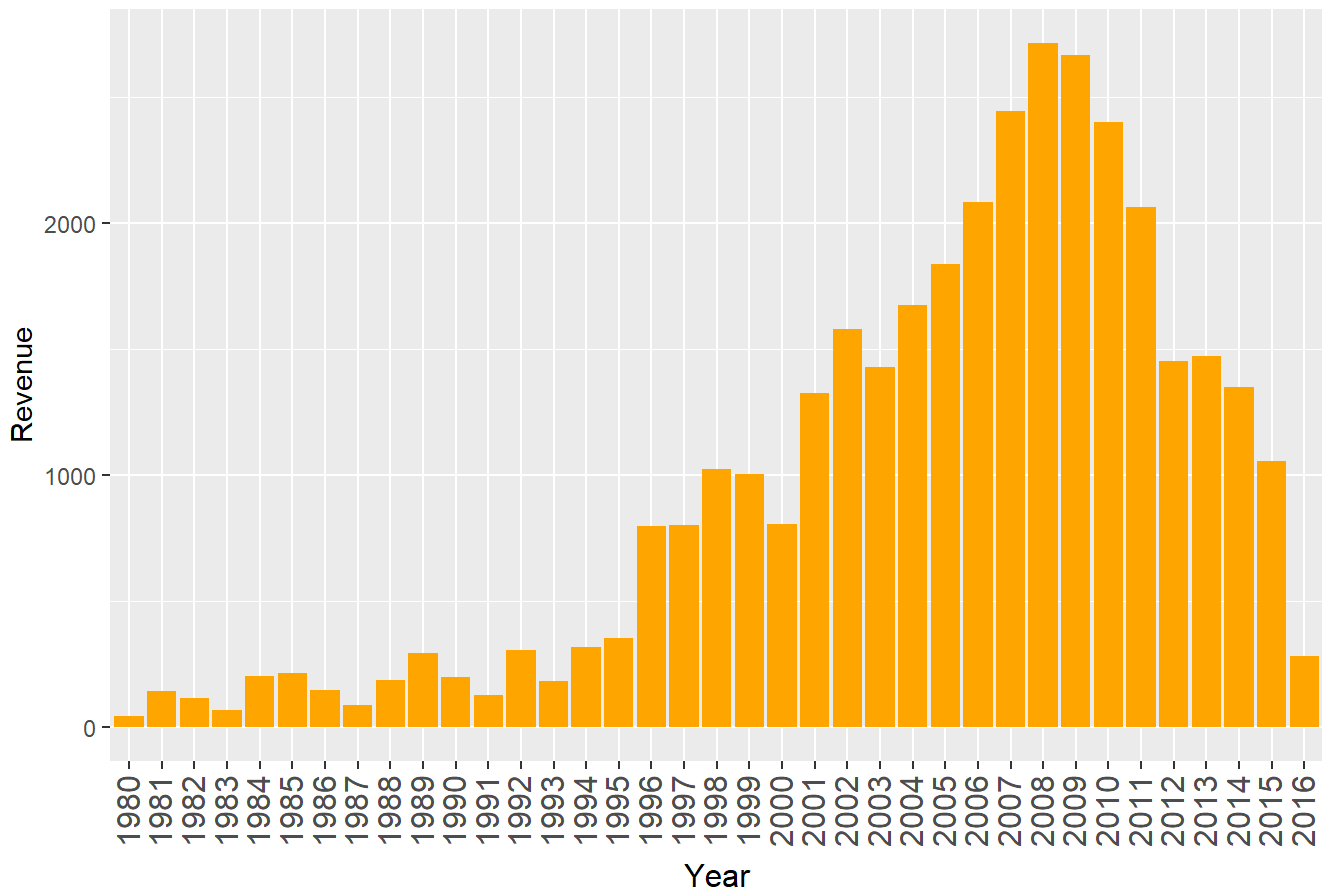
  return(theme(axis.text.x = element_text(size = 12, vjust = 0.4), plot.title = element_text(s
ize = 16, vjust = 2),axis.title.x = element_text(size = 12, vjust = -0.33)))

}
```

```
total_global_revenue_by_year <- VGS %>%
  group_by(Year) %>%
  summarize(Revenue = sum(Global_Sales))

ggplot(total_global_revenue_by_year, aes(Year, Revenue)) +
  geom_bar(fill = "orange", stat = "identity") +
  theme_1() +
  ggtitle("Global Video Game Sales Revenue by Year")
```

Global Video Game Sales Revenue by Year



```
VGS3 <- VGS3[!(VGS3$Year %in% c("N/A", "2017", "2020")),]

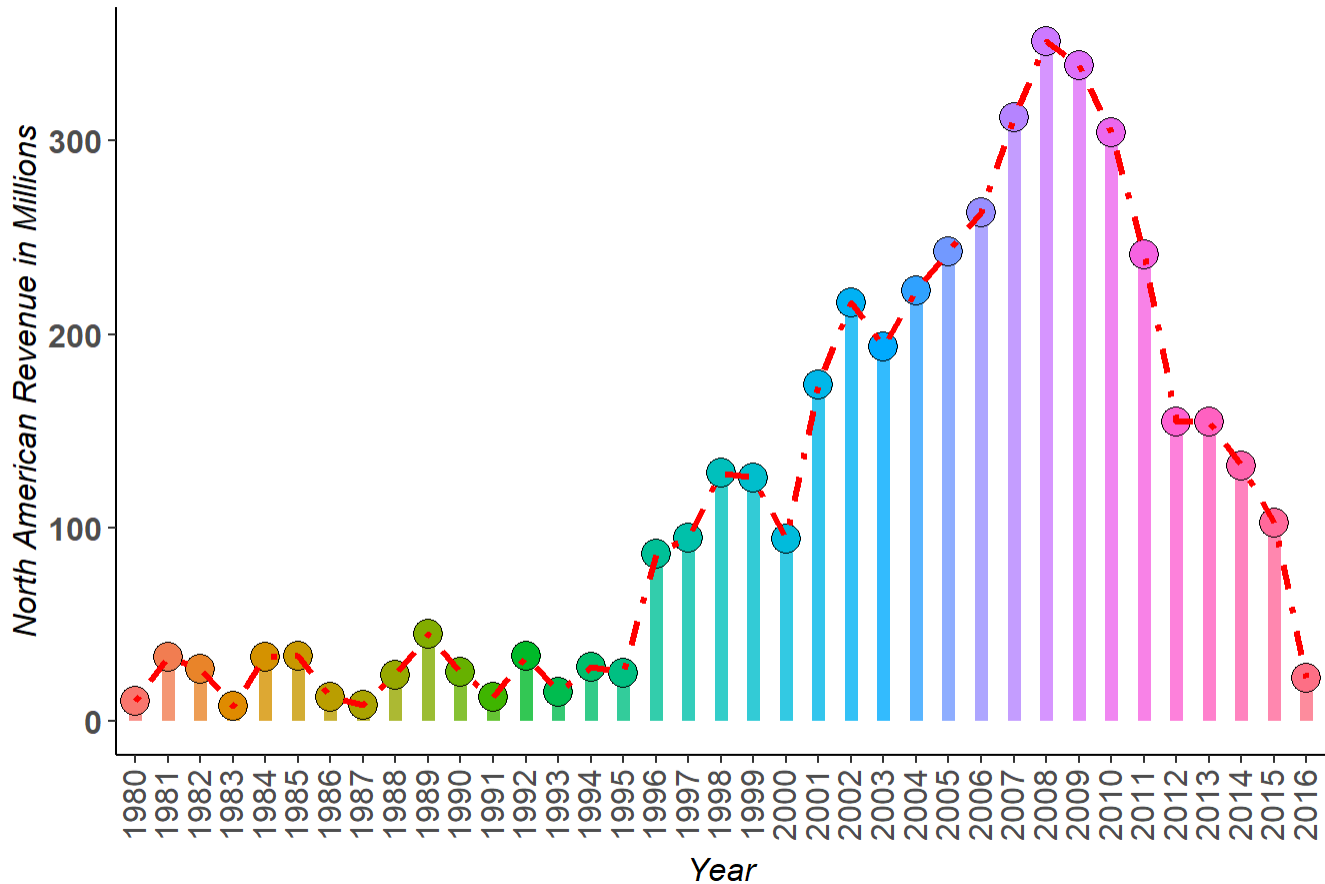
df_NA_Sales <- aggregate(list(NA_Sales = VGS3$NA_Sales), list(Year = VGS3$Year), sum)
df_NA_Sales <- df_NA_Sales[order(df_NA_Sales$NA_Sales), ]

b <- ggplot(data = df_NA_Sales, mapping = aes(x = Year, y = NA_Sales)) +
  geom_segment(aes(xend=Year, yend=0, color = Year), size = 2.3, alpha = .8) +
  geom_point(mapping = aes(fill = Year), size = 5, shape = 21) +
  geom_line(group = 1, size = 1.1, linetype = 10, color = "red") +
  xlab("Year") +
  ylab("North American Revenue in Millions") +
  theme_classic() +
  theme_1() +
  theme(plot.title = element_text(size = 18, face = "bold", hjust = .5),
        axis.title.x = element_text(size = 12, hjust = .5, face = "italic"),
        axis.title.y = element_text(size = 12, hjust = .5, face = "italic"),
        axis.text.x = element_text(size = 12),
        axis.text.y = element_text(size = 12, face = "bold"),
        legend.position = "none") +
  ggtitle("North American Video Game Sales Revenue by Year")
```

```
## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use `linewidth` instead.
```

plot(b)

North American Video Game Sales Revenue by Year



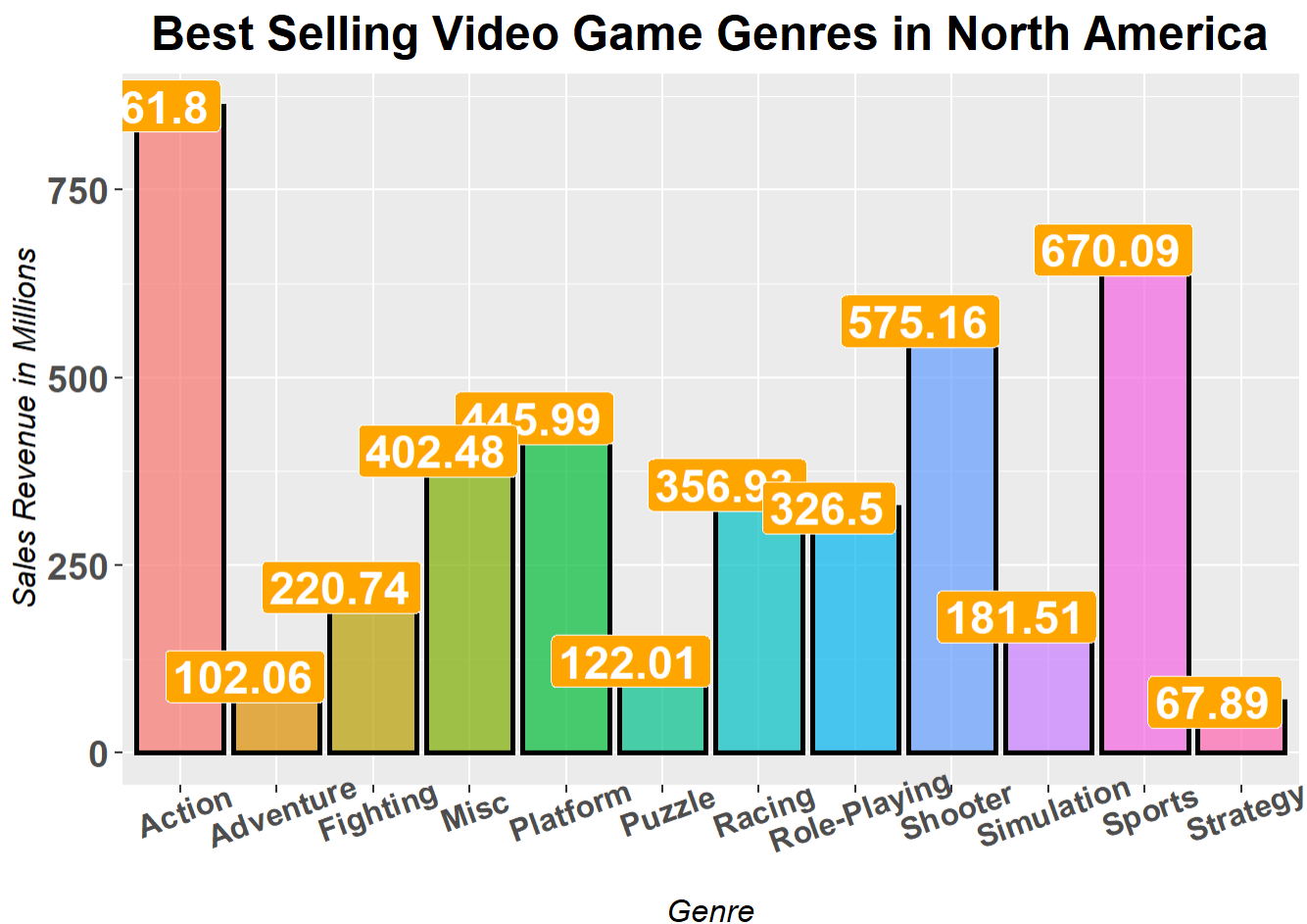
The analysis is clear that North American Sales account for the highest grossing sales (therefore it is important to understand the market demands in the specific region for our video games sales. Additionally, the data has also been cleaned by removing the incomplete data with missing variables that identify the Year of popularity for North American Sales. However, this data could have a minimal impact as the first question of the analysis only looked at the highest grossing years with revenues more than \$100 Million. According to the data between 2004 and 2007, North American video games sales saw their highest revenues with increasing sales finally reaching their peaked in 2008 when they achieved \$351.44 Million in total revenue. In the years following their peak in 2008, revenues from sales remained above \$100 Million until 2015.

```

g_name_NA <- aggregate(list(NA_Sales = VGS2$NA_Sales), list(Genre = VGS2$Genre), sum)
g_name_NA <- g_name_NA[order(g_name_NA$NA_Sales, decreasing = T), ]

options(repr.plot.width = 14, repr.plot.height = 6)
ggplot(data = g_name_NA, mapping = aes(x = Genre, y = NA_Sales)) +
  geom_bar(stat = "identity", mapping = aes(fill = Genre), alpha = .7, size = 1, color = "black") +
  geom_label(mapping = aes(label=NA_Sales), fill = "orange", size = 6, color = "white", fontface = "bold", hjust=.7) +
  ggtitle("Best Selling Video Game Genres in North America") +
  xlab("Genre") +
  ylab("Sales Revenue in Millions") +
  theme(
    plot.title = element_text(size = 18, hjust = .5, face = "bold"),
    axis.title.x = element_text(size = 12, hjust = .5, face = "italic"),
    axis.title.y = element_text(size = 12, hjust = .5, face = "italic"),
    axis.text.x = element_text(size = 12, face = "bold", angle = 20),
    axis.text.y = element_text(size = 14, face = "bold"),
    legend.position = "none")

```

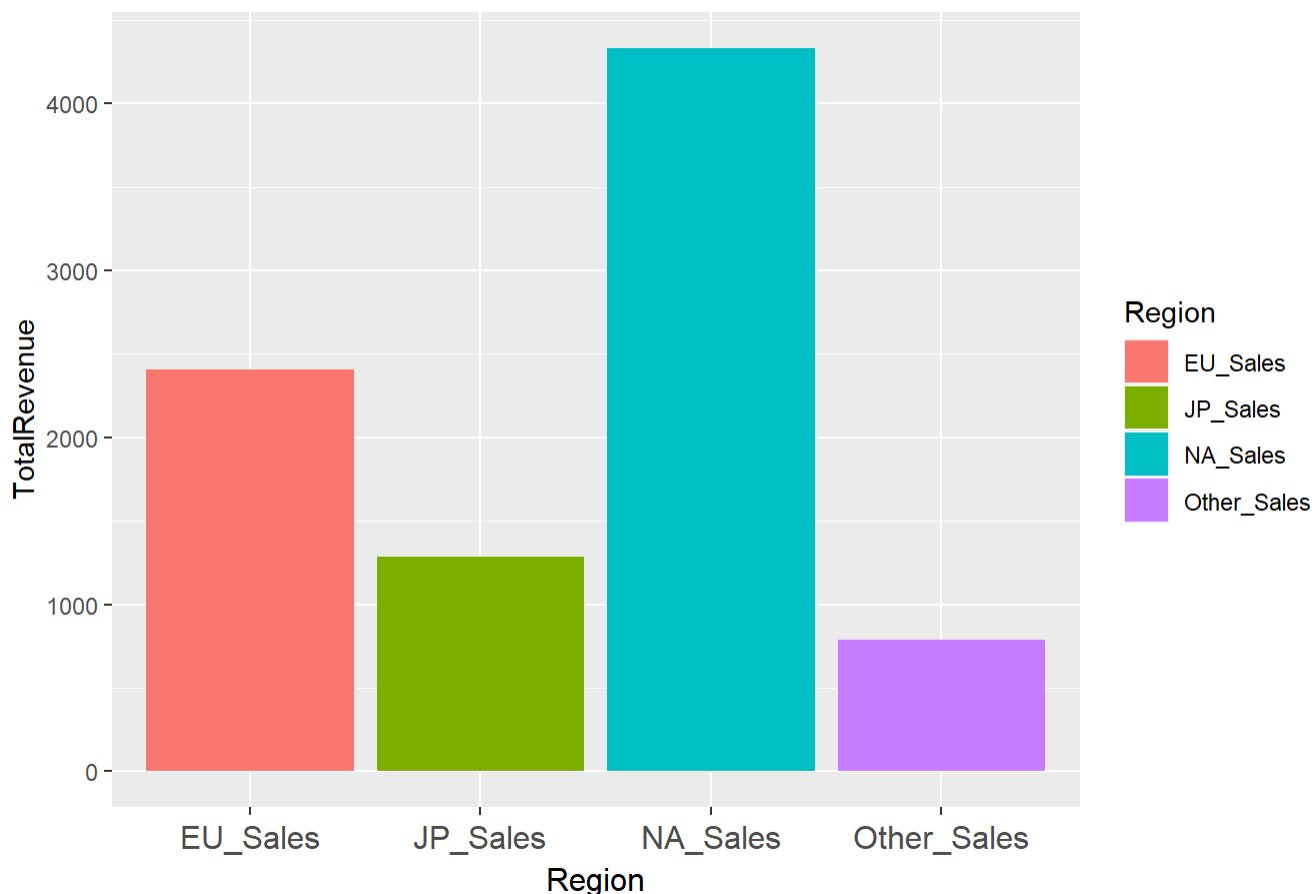


To summarize the findings of North American sales figures, we see that genre of Action player video games dominated sales for the years 2007 to 2014. Additionally, sports genre video games dominated revenue sales figures in 2006 and kept slightly lower total revenue and remained competitive in the video game marketplace.

Finally, action genre video game sales started a steady decline in sales beginning in 2013 and continuing into 2016. In the 2015 the shooter style videos games reach peak sales revenue of \$30 Million and beating both action and sports genre video games. For the year in which peak sales were the highest in North America the genre of video games that were the highest selling involved the Action genre, followed by Sports and shooter style games. However, Action genre video games have beat out the Sports video games sales in North America since 2007, when they overtook the sales revenue from the previously held sports genre in 2006. Additionally, the data set indicates that Action style video games maintain their level for highest grossing sales until 2014, which is when they lost their edge to shooter style video games. Analysis below we look at total revenue by regions outside the North American Marketplace.

```
by_regions <- VGS %>%  
  group_by(Region) %>%  
  summarize(TotalRevenue = sum(Revenue), Percentage = TotalRevenue/sum(VGS$Revenue) * 100) %>%  
  arrange(desc(TotalRevenue))  
  
ggplot(by_regions, aes(Region, TotalRevenue, fill = Region)) +  
  geom_bar(stat = "identity") +  
  theme_2() +  
  ggtitle("Total Revenue by Region") +  
  theme(legend.position = "right")
```

Total Revenue by Region

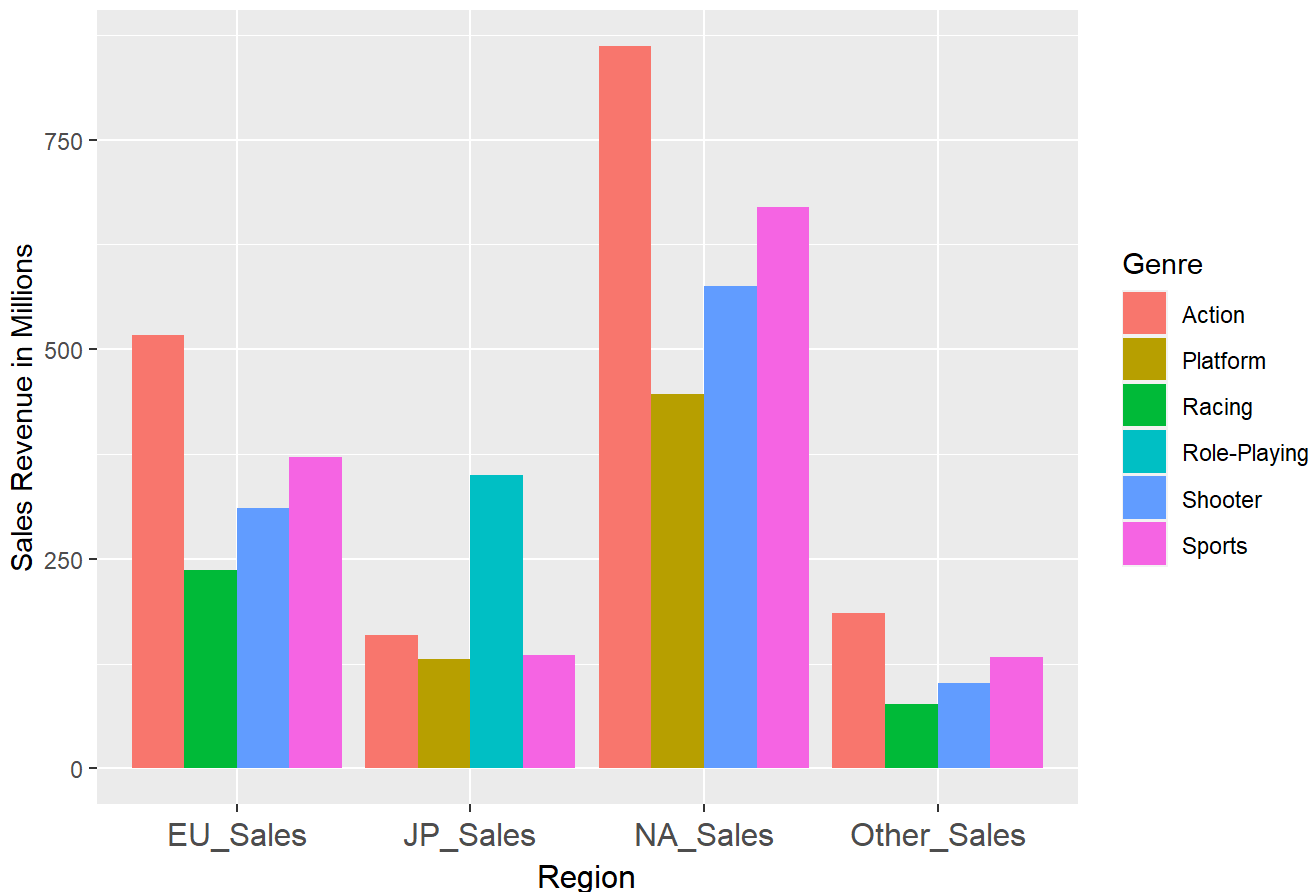


```
top_genres_region <- VGS %>%
  group_by(Region, Genre) %>%
  summarize(Revenue = sum(Revenue)) %>%
  arrange(desc(Revenue)) %>%
  top_n(4)
```

```
## `summarise()` has grouped output by 'Region'. You can override using the
## `.groups` argument.
## Selecting by Revenue
```

```
ggplot(top_genres_region, aes(Region, Revenue, fill = Genre)) +
  geom_bar(position = "dodge", stat = "identity") +
  ggtitle("Top 4 Genres by Sales Revenue in each Region") +
  ylab("Sales Revenue in Millions") +
  xlab("Region") +
  theme_2() +
  theme(legend.position = "right")
```

Top 4 Genres by Sales Revenue in each Region



When looking at the sales data in countries outside of the North American marketplace, we see a similar preference for (Action, Sports, and Shooter) genre games in European countries. However, in the Japanese marketplace we see that Role-playing video games are the best selling and shooter style video games are the lowest selling. Finally, in other regions action and sports and shooter style games are the highest selling.

Methodology:

My methodology for selecting and the video game dataset was to better understand what genre of video games our popular and on a a global scale over a number of years. Additionally,, I found that my research benefits us because we can potentially lesson the ethical dilemma over saturation the marketplace with shooter based video games. However, the con imposed on controlling this variable is the potential for less revenues and bias in our companies' final decisions. Finally, further research could benefit us as the recent data may also suggest that shooter style games may have already been declining after the 2016 based on the potential limitations of this data set.

Conclusion:

I believe that market specific video game genres such as shooter style may have reached peak saturation. With this data I believe it would be wise to potentially avoid this market space entirely due to negative revenue stream generation and limited growth potential; as well as the ethical dilemmas surrounding violence among young individuals. However, because of the popularity of sports style games in the European market and North American Market, we should focus our efforts on capturing more of marketplace in the future. Additionally, the research is a little surprising to me because they seem to differ from the video games I enjoyed and played with my family and friends. Such games I remember playing as a child include Super Mario Bros, Tetris, and Mario Kart, which were all very important games to own as a child. Finnaly, it appear that today's gamer's in North American and European have a preference for Action, Sports and Shooter style video games. Overall, limiting the data research to North American and European regions may not paint the entire picture of popularity. The Japanese marketplace sales of Role Playing games data resembles the sales revenue of Sports and shooter revenues in the North American and European Market place.

Future Analysis:

I would like to extend my analysis further once we acquire updated and more recent (2017 thru 2023) video game statistics as they become available. By obtaining additional and more current data we can make stronger conclusions that we have missed with this analysis. However, with current data we can more accurately forecast what potential sales revenue patterns could look like for our business in the years to come. This data can help us achieve financial benchmarks across any changing customer preferences in video gaming purchases, platforms, and geographic targets.