Video Game Industry: Competitive Landscape Analysis

Joshua McCleary

Nicholas McIntosh

Jennifer M. Rogers

James M. Watson

Syracuse University

**Introduction**

Sales of video games have exponentially increased since their introduction to the mainstream in the 1970s. This growth is expected to continue in 2020 with a projected sales volume across the globe totaling $90 billion (WePC, 2020). What started as an industry of one or two games, blossomed into an industry with multiple genre’s, by multiple publishers, on numerous platforms. Our research and analytics drove us to answer the question: Based on previous sales for the top Publisher could we make a recommendation of which categorical Genre they create a game for?

In order to answer this question, we sought out answers to smaller questions that would efficiently assist in helping us to predict this outcome. First, we looked at the Publisher and Platform performance over time. Second, we looked at the rank of Publishers based on total market share. Third, we sought to find the rank of each genre.

**Video Game Sales Dataset**

The file used for our exploratory analysis was compiled by Halim Noor in March of 2020 and was pulled from Kaggle for this project. Below are the features of this data set.

|  |  |
| --- | --- |
| Column Name | Description |
| Name | This contains the name of the individual game. |
| Platform | This is the platform in which the individual game is played on. There is a total of 31 unique gaming platforms within this data set. |
| Year of Release | This column contains data starting in 1980 and ending in 2020. |
| Genre | This contains the categorical genre of individual games. There are 12 unique categorical values of genre. |
| Publisher | This contains the publishers of each individual game. There are 579 unique publishers. |
| NA\_Sales | This lists the sales of each individual video game in North America. |
| EU\_Sales | This lists the sales of each individual video game in Europe. |
| JP\_Sales | This lists the sales of each individual video game in Japan. |
| Other\_Sales | This lists the sales of each individual game in an unspecified market(s). |
| Global\_Sales | This lists the total sales of each individual game within the combined markets. |

Table 1: Features of Data

**Data Cleansing and Management**

After loading the file into “R Studio”, we did some initial data cleansing. The first step was to manage N/A’s in the data frame, we did this by doing “vgsales<-na.omit(vgsales)”. With the release of the Wii system, Wii Sports became the highest-ranking game and outsold all others by a substantial amount. The substantial difference in sales are inflated due to Wii Sports being sold in a “bundle”; meaning, this game was sold with the Wii Console, Wii Controllers, or other merchandise. The overall sales for the game could not be accounted for on a standalone basis. After consideration and implementation of methods to manage the skew, we deleted Wii Sports to get a greater understanding of the overall picture “VG<-VG[-c(1),]”. With the removal of Wii Sports, which was initially in the rank position of “1”, we renumbered the rank to account for this missing data element “VG$Rank<-c(1:1697)”.

Each team member created their own portion of individual coding that contributed to answering our main business question. Within the coding, we parsed the data to pull the elements needed for our individual analysis. Below is a portion of the unexecuted code:

* To run an analysis of publishers based on market sales simplified data frames were created for Global\_Sales, NA\_Sales, EU\_Sales, and JP\_Sales with a final data frame called “pubmerge” created to house these smaller data frames:

Pub1<-tapply(VG$Global\_Sales,VG$Publisher,sum)

Pub2 <- cbind(rownames(Pub1), data.frame(Pub1, row.names=NULL))

names(Pub2)[names(Pub2)=="rownames(Pub1)"]<-"Publisher"

names(Pub2)[names(Pub2)=="Pub1"]<-"Global Sales"

Pub3<-tapply(VG$NA\_Sales,VG$Publisher,sum)

Pub3<-cbind(rownames(Pub3),data.frame(Pub3,row.names=NULL))

names(Pub3)[names(Pub3)=="rownames(Pub3)"]<-"Publisher"

names(Pub3)[names(Pub3)=="Pub3"]<-"North America Sales"

Pub4<-tapply(VG$EU\_Sales,VG$Publisher,sum)

Pub4<-cbind(rownames(Pub4),data.frame(Pub4,row.names=NULL))

names(Pub4)[names(Pub4)=="rownames(Pub4)"]<-"Publisher"

names(Pub4)[names(Pub4)=="Pub4"]<-"Europe Sales"

Pub5<-tapply(VG$JP\_Sales,VG$Publisher,sum)

Pub5<-cbind(rownames(Pub5),data.frame(Pub5,row.names=NULL))

names(Pub5)[names(Pub5)=="rownames(Pub5)"]<-"Publisher"

names(Pub5)[names(Pub5)=="Pub5"]<-"Japan Sales"

pubmerge<-data.frame(c(Pub2,Pub3,Pub4,Pub5))

* A simplified data frame was created based on the “head” of each publishers (pubmerge 2.2, etc.) after placing into ascending order (pubmerge2, etc.) to isolate and determine the top three publishers across the North America, Europe, Japan, and globally. This was completed using the melt function and producing “pubmelt2”:

“pubmerge2<-pubmerge[order(-pubmerge$Global.Sales,pubmerge$Publisher),]

pubmerge3<-pubmerge[order(-pubmerge$North.America.Sales,pubmerge$Publisher),]

pubmerge4<-pubmerge[order(-pubmerge$Europe.Sales,pubmerge$Publisher),]

pubmerge5<-pubmerge[order(-pubmerge$Japan.Sales,pubmerge$Publisher),]

pubmerge2.2<-head(pubmerge2)

pubmerge3.3<-head(pubmerge3)

pubmerge4.4<-head(pubmerge4)

pubmerge5.5<-head(pubmerge5)

puballup<-data.frame(pubmerge2.2,pubmerge3.3,pubmerge4.4,pubmerge5.5)

pubmelt<-melt(puballup,"Publisher")

pubmelt2<-pubmelt[-c(7:12),]”

* A simplified data frame was created run an analysis of market sales based on Genre which were combined into the table “genreSalesTotal”. Global sales were left alone in the simplified table “globalSales”:

“globalSales<-tapply(VG$Global\_Sales,VG$Genre,sum)

genre<-rownames(globalSales)

GenreSalesGlobal<-data.frame(genre,globalSales)

naSales<-tapply(VG$NA\_Sales,VG$Genre,sum)

euSales<-tapply(VG$EU\_Sales,VG$Genre,sum)

jpSales<-tapply(VG$JP\_Sales,VG$Genre,sum)

otherSales<-tapply(VG$Other\_Sales,VG$Genre,sum)

genreSalesTotal<-data.frame(naSales,euSales,jpSales,otherSales)”

* In order to create a bar plot that would give an accurate visual representation of genre performance in each market, a data frame was created to show which genres are most popular. This was backed by market sales for those genre’s:

“genreSalesTotal$genre<-rownames(genreSalesTotal)

rownames(genreSalesTotal)<-NULL

market<- c(rep("NaSales" , 12) , rep("EUSales" , 12) , rep("JPSales" , 12) , rep("OtherSales" , 12) )

genre1 <- rep(c("Action" , "Adventure","Fighting", "Misc", "Platform","Puzzle","Racing","Role-Playing",

"Shooter", "Simulation","sports", "Strategy") , 4)

Unit <- c(877.83,105.80,223.59,410.24,447.05,123.78,359.42,327.28,582.60,183.31,683.35,68.70,525.00, 64.13,101.32,215.98,201.63,50.78,238.39,188.06,313.27,113.38,376.85,45.34,159.95,52.07,87.35,107.76,130.77,57.31,56.69,352.31,38.28,63.70,135.37,49.46,187.38,16.81,36.68,75.32,51.59,12.55,77.27,59.61,102.69,31.52,134.97,11.36)

data <- data.frame(market,genre1,Unit)”

* In order to determine the genre with the most sales across markets a separate data frame was completed. Each individual variable of sales were given their own separate data frames. They were then merged by Genre into “genreallup”.

sbggs<-sqldf("SELECT sum(Global\_Sales), Genre FROM VG group by Genre")

sbgnas<-sqldf("Select sum(NA\_Sales), Genre FROM VG group by Genre ")

sbgeus<-sqldf("Select sum(EU\_Sales), Genre FROM VG group by Genre")

sbgjps<-sqldf("SELECT sum (JP\_Sales), Genre FROM VG group by Genre")

sbgos<-sqldf("Select sum(Other\_sales), Genre FROM VG group by Genre")

genreallup<-merge(sbggs,sbgnas,"Genre")

genreallup<-merge(genreallup,sbgeus,"Genre")

genreallup<-merge(genreallup,sbgjps,"Genre")

genreallup<-merge(genreallup,sbgos,"Genre")

colnames(genreallup)<-c("Genre","Global\_Sales", "NA\_Sales", "EU\_Sales","JP\_Sales","Other\_Sales")

**Data Analysis Methods**

Several statistical methodologies were implemented when analyzing the data. Each individual member contributed a significant amount of statistical analysis in order to assist in determining the top publisher, genre, platform, rank, and sales. Using sqldf functionality in R Studio, individual tables were made to focus on the “Platform and Publisher Performance”. This helped to manipulate the data in order to identify the top Platform and Publisher overall. Visualizations were created to support the outcome. The following statistical methodology was implemented to determine the spread of sales among publishers: standard deviation (sd) of each individual sales market, the summary function of each individual publisher to determine basic information concerning the data, and linear modeling (lm) to determine how the overall publisher sales influence or predict the global sales.

In the section “Rank and Sales of Genre”, team member created a function that calculates the mean, median, min, max, sd, and quantile at the .05 and .95 markers. This has helped to identify the central tendencies and spread of the data as it relates to each genre (Saltz & Stanton, 2018, p. 145). Team members also utilized linear modeling in order to tell the story of how each individual sales market influenced the dependent variable of global sales. This output assisted in determining the different amount of sales per market from each genre and allows us to see how the sales per genre can assist in predicting the global sales.

**Analysis Summaries**

Publisher and Platform Performance

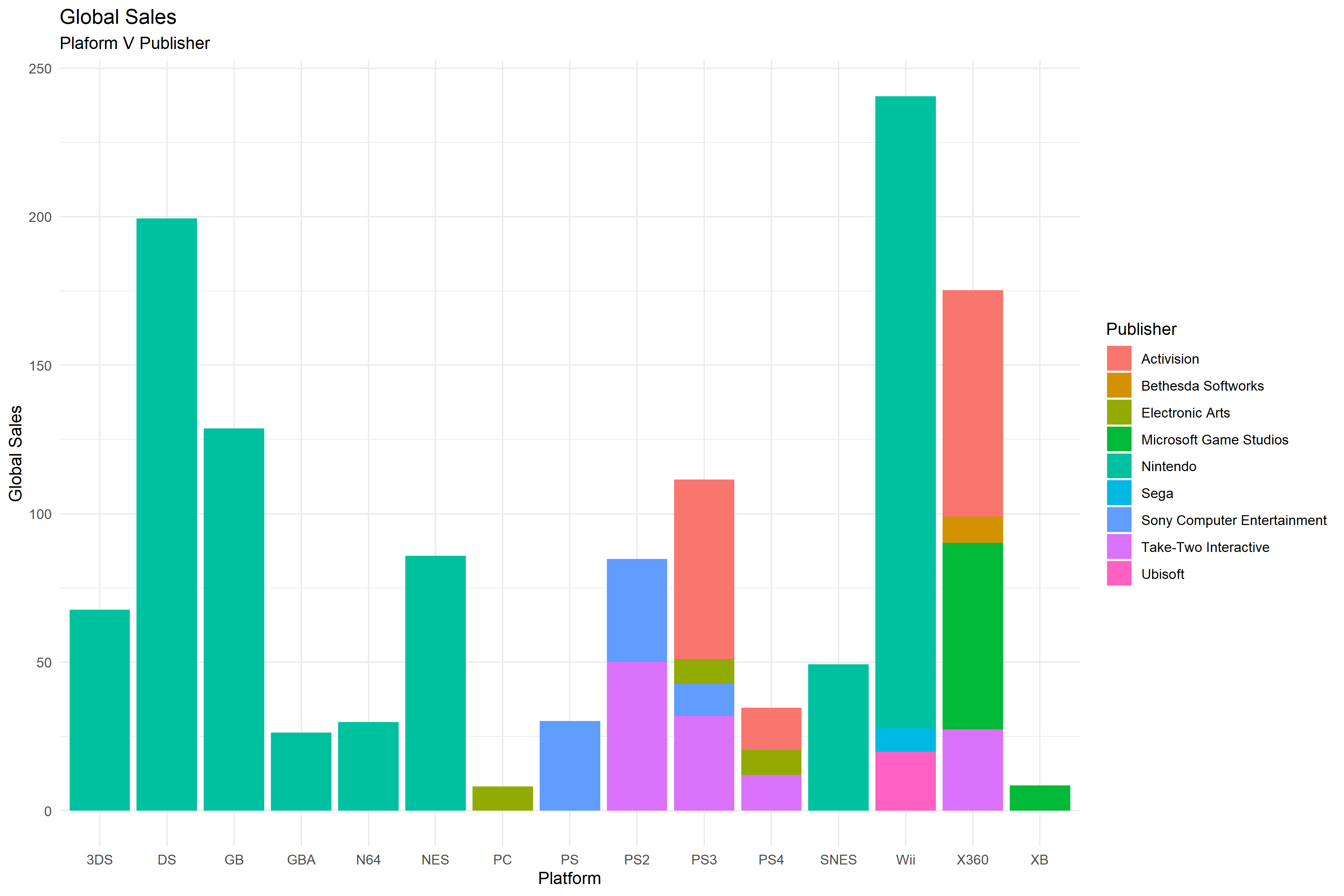
Video games were introduced to homes on August 28, 1972, with the introduction of the Magnavox Odyssey (Cifaldi, 2012). The idea was certainly the first of its kind to be made commercially available to at home audience and paved the way for numerous other platforms to enter the market (Cifaldi, 2012). In today’s day and age people now have a choice of which console, or platform, they choose to play their video games on. This makes the market incredibly competitive.

In order to determine the top three platforms based on sales, individual data tables were created using the sqldf function (JPplatform<-sqldf("Select sum(JP\_Sales), Platform FROM VG group by Platform"). Once the individual data frames were completed, they were then sorted into data frames with ascending order based on sales (JPplat1<-JPplatform[order(-JPplatform$JP.Sales,JPplatform$Platform),]. Finally, using the “head” function, the top six of each data frames were then merged into the data frame “Platallup1”(Platallup<-data.frame(Gplat1,NAplat1,EUplat1,JPplat1), which then displayed the results in Table 2 below.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Rank** | **Platform (Global)** | **Global Sales** | **Platform (NA)** | **NA Sales** | **Platform (EU)** | **EU Sales** | **Platform (JP)** | **JP Sales** |
| **1** | **PS2** | **1255.77** | **X360** | **601.05** | **PS3** | **343.71** | **DS** | **175.57** |
| **2** | **X360** | **979.6** | **PS2** | **583.84** | **PS2** | **339.29** | **PS** | **139.82** |
| **3** | **PS3** | **957.89** | **Wii** | **466.22** | **X360** | **280.58** | **PS2** | **139.2** |

Table 2: Platform Rank by Market-Top 3 Only

Claiming a spot in the top three in all markets is the PS2. This comes as no surprise, in 2017 the PlayStation2 console became the most popular console in the world with 157.68 million units sold (WePC, 2020). Our platform performance suggests that Sony is a top contender for Platform to use when publishing a game. With this in mind, understanding how the publisher impacts sales is also imperative to understanding how to manipulate the consumers within an individual market.



VG %>%

filter(Global\_Sales >= 8 & Global\_Sales <= 40.5) %>%

ggplot() +

aes(x = Platform, fill = Publisher, weight = Global\_Sales) +

geom\_bar() +

scale\_fill\_hue() +

labs(y = "Global Sales", title = "Global Sales ", subtitle = "Plaform V Publisher") +

theme\_minimal()

Figure 1: Publisher and Platform performance

To better understand Platform’s sales, standard deviation was used to get a better understanding of how wide the data is spread. Currently, the sd for North America takes the top spot with the largest spread at 175.28. Europe and Japans sd are placed second and third respectively with an overall sd of 104.52 and 51.28 respectively. To better understand how the sales of Platform specifically relate to the global sales, linear modeling was implemented and created for each of the individual markets.

North America (P1 <- lm(formula = NA.Sales ~ Global.Sales, data=Platallup), had the lowest p-value, with a rather statistically significant out at 2.2e-16. The Adjusted R-Squared output also showed great statistical significance at 98.17%. Based on the output, North America’s sales have the greatest influence on Global Sales. We can infer, that if North America has a successful year within the video game industry, then globally, we can expect the impact to correlate. Slightly lower than North America is the European and Japanese sales of Publisher. With an Adjusted R-Squared of 96.29% and 91.19% respectively.

Understanding how the market has changed over time is also helpful in understanding the top Platform performance, as well as Publisher performance. The visual below allows us to see how the share of Publisher and Platform has changed over time. This also allows us to see when and where the video game industry really became a pertinent part of the global market for consumer spend:

VG %>%

filter(Global\_Sales >= 9 & Global\_Sales <= 40.5) %>%

ggplot() +

aes(x = Publisher, fill = Platform, weight = Global\_Sales) +

geom\_bar() +

scale\_fill\_hue() +

labs(y = "Global Sales") +

coord\_flip() +

theme\_minimal() +

facet\_wrap(vars(Year))

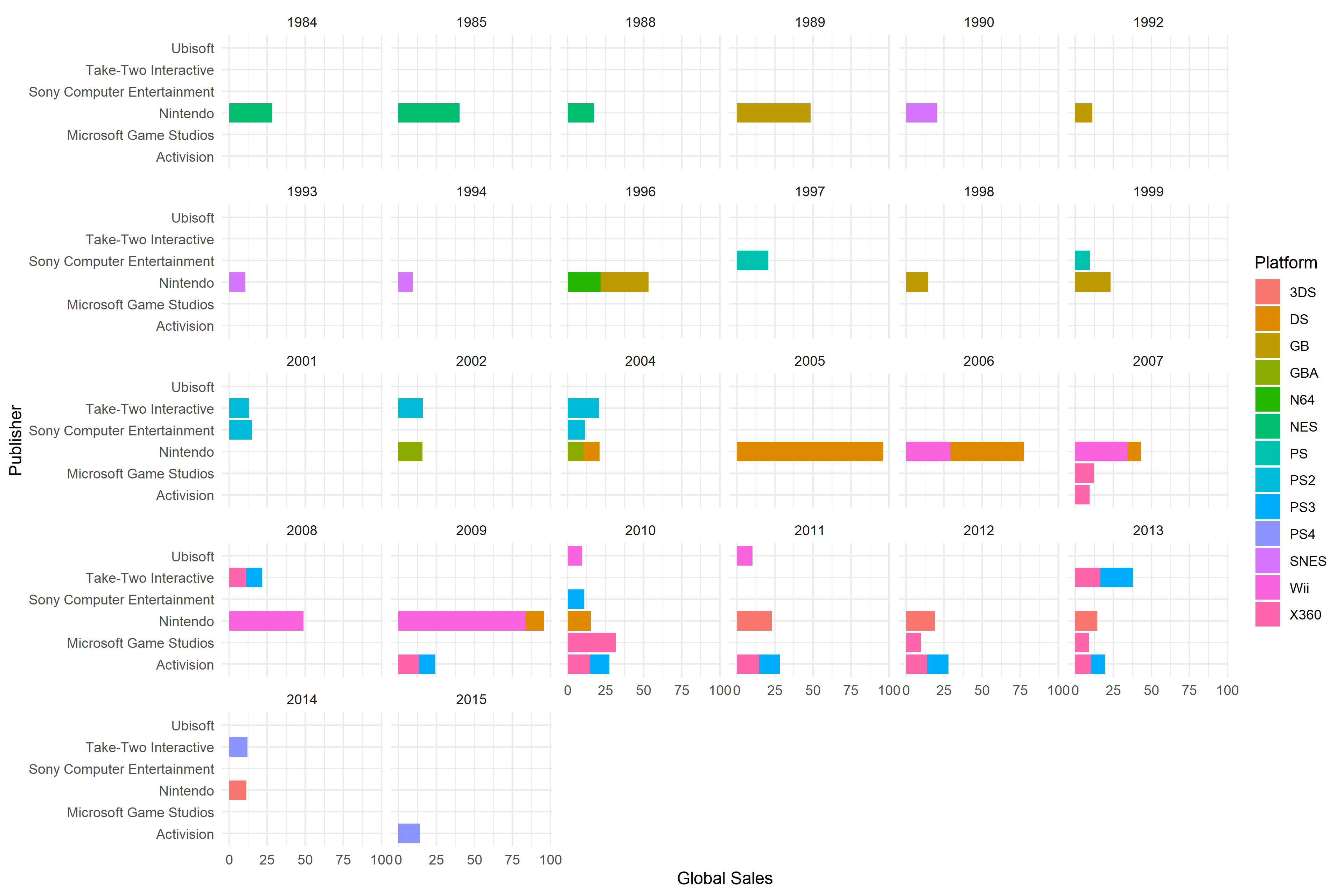
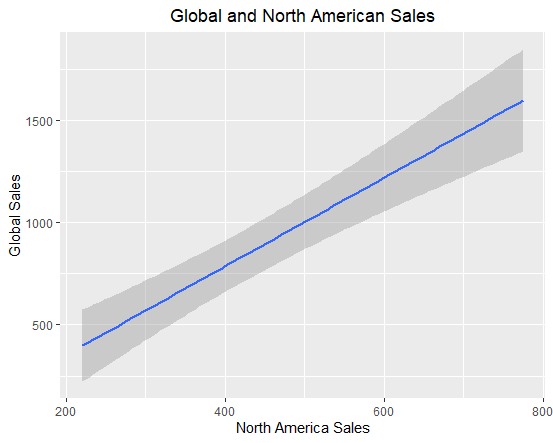


Figure 2: Publisher and Platform Performance Over Time

To further understand how the rank of each publisher correlates to the overall market share compared to global sales linear modeling (lm) was employed to assist in answering these questions. Furthermore, the use of standard deviation to assist in understanding the overall spread of the data was used. Below are graphics and results with explanations of the statistical analysis as well as conclusions drawn from the analysis.



**Based on Publisher by Market**

“lm(formula = puballup$North.America.Sales ~ puballup$Global.Sales,

data = puballup)”

“Residual standard error: 49.94 on 4 degrees of freedom

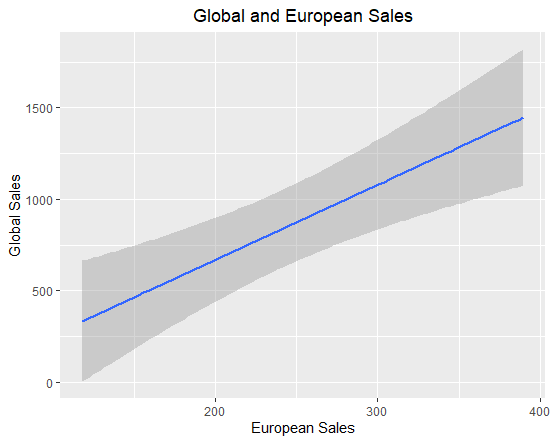
Multiple R-squared: 0.9599,

Adjusted R-squared: 0.9498

F-statistic: 95.65 on 1 and 4 DF,

p-value: 0.0006125”

Figure 3



**Based on Publisher by Market**

“lm(formula = puballup$Europe.Sales ~ puballup$Global.Sales, data = puballup)”

“Residual standard error: 42.96 on 4 degrees of freedom

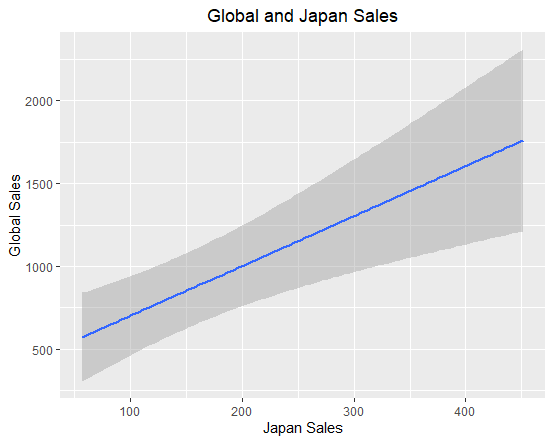
Multiple R-squared: 0.8841,

Adjusted R-squared: 0.8551

F-statistic: 30.51 on 1 and 4 DF,

p-value: 0.005245”

Figure 4



**Based on Publisher by Market**

“lm(formula = puballup$Japan.Sales ~ puballup$Global.Sales, data = puballup)”

“Residual standard error: 61.83 on 4 degrees of freedom

Multiple R-squared: 0.8681,

Adjusted R-squared: 0.8351

F-statistic: 26.32 on 1 and 4 DF,

p-value: 0.006837”

Figure 5

From the data, we can see that North American Sales and Global Sales have the highest coefficient of determination, at 94.98%. Essentially, North America Sales has the greatest influence on our dependent variable of Global Sales. We can assume, that future forecasting would be weighted towards what the expected, or predicted, North American Sales were for a particular year or decade. European sales and Japanese sales have a coefficient of determination of 85.56% and 83.51% respectively.

Calculating the spread of these variables using standard deviation, sd(puballup$Europe.Sales) and sd(puballup$Japan.Sales), their measure of dispersion is relatively wide with a sd of 112.88 and 152.26 respectively. Comparatively, North America Sales has a sd (sd(puballup$North.America.Sales) of 222.93. The greatness of spread also helps to visualize how vast the sales in the markets distributed by the individual publishers. The top three publishers based on the individual sales market, as well as global sales, is listed in “Table 3” (puballup<-data.frame(pubmerge2.2,pubmerge3.3,pubmerge4.4,pubmerge5.5).

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Rank** | **Publisher** | **Global Sales** | **Publisher** | **North America Sales** | **Publisher** | **Europe Sales** | **Publisher** | **Japan Sales** |
| #1 | Nintendo | 1703.62 | Nintendo | 775.38 | Nintendo | 389.72 | Nintendo | 451.65 |
| #2 | Electronic Arts | 1110.15 | Electronic Arts | 595.07 | Electronic Arts | 371.27 | Namco Bandai Games | 127.07 |
| #3 | Activision | 727.11 | Activision | 429.7 | Activision | 215.53 | Konami Digital Entertainment | 91.3 |

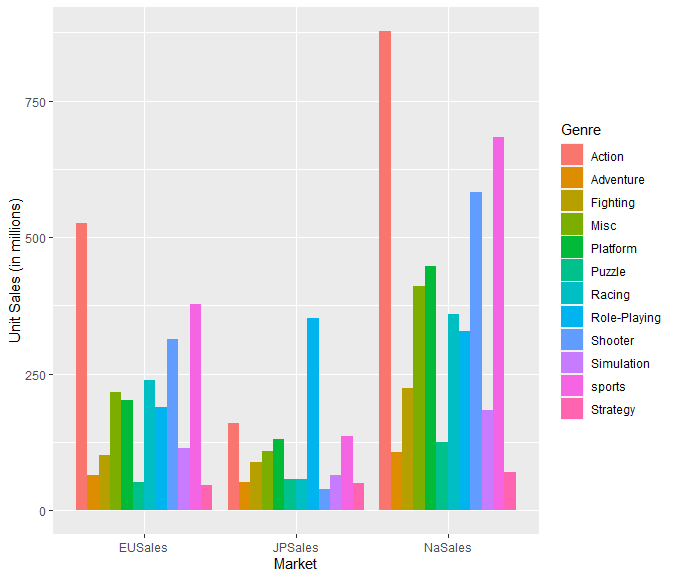
Table 3: Publisher Rank by Market

Rank and Sales of Genre

Examining the rank and sales of genre will continue the story that the data is trying to tell us. Video games are relatively recent in the history of this world, the “Video Game Sales” data set by Halim Noor (2020) has data that starts in 1980 and ends in 2020. Video game consoles first became available for in home use in 1972, with the launch of the Magnavox Odyssey (Willeart, 2018). In order to understand the full picture of sales, it is necessary to peruse the data at a micro level.

The function “statsVG” was created in order to run measures of central tendency and measures of dispersion. Central tendency measures include mean and mode. Measures of dispersion include standard deviation, range, and quantile functions. This assists in understanding how the data differentiates itself from across markets. Output such as range can tell us about how wide the data is spread out, while mean gives us the average of all data.

North America sales had the highest mean of sales at 362.62, as well as the greatest range within their data, with a minimum of 68.7 and a maximum of 877.83. In Figure 4, we can see that in North America, the highest sales were in the “Action” genre, and the lowest sales are allocated to Strategy games. In Europe, while the sales are less than North America with a max sales of 525 and a minimum sales at 45.34, in Figure 4 it is easily identifiable that European sales follow a similar path as North America’s sales with the “Action” genre taking the top spot, and the “Strategy” genre taking the bottom spot.



statsVGstatsVG(genreSalesTotal$naSales)

mean: 362.6217 median: 343.35 min: 68.7 max 877.83 sd: 245.5669 quantile(0.05-0.95): 89.105 -- 748.046

statsVG(genreSalesTotal$euSales)

mean: 200.4258 median: 194.845 min: 45.34 max 525 sd: 142.7231 quantile(0.05-0.95): 48.332 -- 427.5565

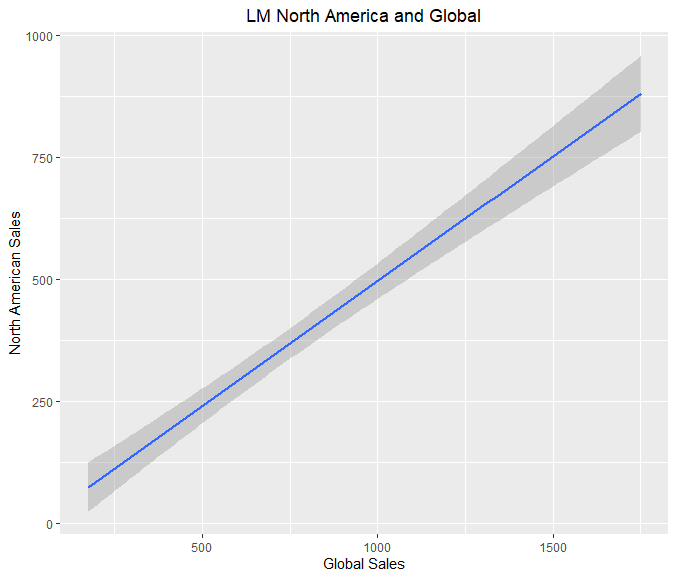
statsVG(genreSalesTotal$jpSales)

mean: 107.2708 median: 75.525 min: 38.28 max 352.31 sd: 86.58279 quantile(0.05-0.95): 44.429 -- 246.512

Figure 6

Japan’s sales account for the lowest amount in the overall global sales. With a minimum of 38.28 and a maximum of 352.31, their sales are much below that of the North America Sales. Interestingly, there is a break in the trend of top genre’s in Japans Market’s. The genre “Role-Playing” takes the top spot at 352.31 in sales. While the genre “Shooter” takes last place with only 38.28 sales. In Figure 6, the genre “Shooter” comes in third in both the European and North American markets.

Linear modeling assists in determining the outcome of genre sales in the individual markets as it relates to the global sales. This helps to give us a better understanding of how the impact of each of these areas directly influence global sales. Furthermore, while linear modeling isn’t necessarily supposed to be used to compare variables in which the summation is included in the dependent variable, understanding the spread of the data and how one correlates to each other is helpful in gaining insight into which market with which genre would be best utilized.



lm(formula = NA\_Sales ~ Global\_Sales, data = genreallup)

Residual standard error: 49.08 on 10 degrees of freedom

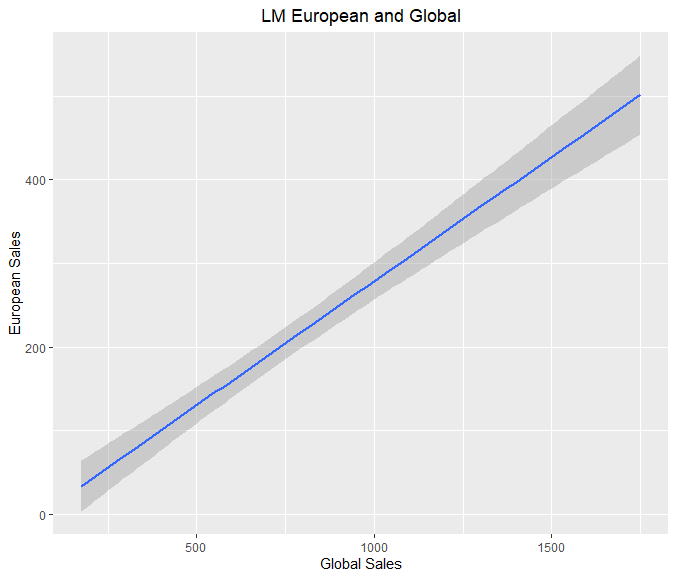
Multiple R-squared: 0.9637,

Adjusted R-squared: 0.9601

F-statistic: 265.4 on 1 and 10 DF

p-value: 1.579e-08

Figure 7



lm(formula = EU\_Sales ~ Global\_Sales, data = genreallup)

Residual standard error: 29.85 on 10 degrees of freedom

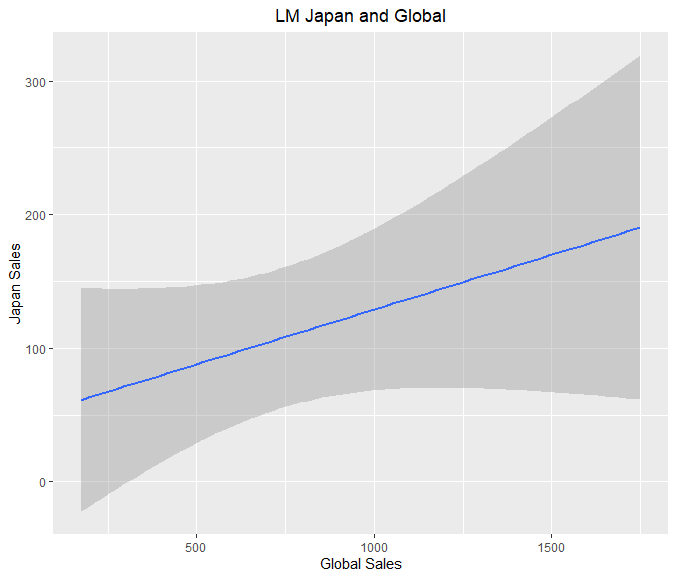
Multiple R-squared: 0.9602,

Adjusted R-squared: 0.9563

F-statistic: 241.5 on 1 and 10 DF

p-value: 2.486e-08

Figure 8



lm(formula = JP\_Sales ~ Global\_Sales, data = genreallup)

Residual standard error: 81.31 on 10 degrees of freedom

Multiple R-squared: 0.1983,

Adjusted R-squared: 0.1182

F-statistic: 2.474 on 1 and 10 DF

p-value: 0.1468

Figure 9

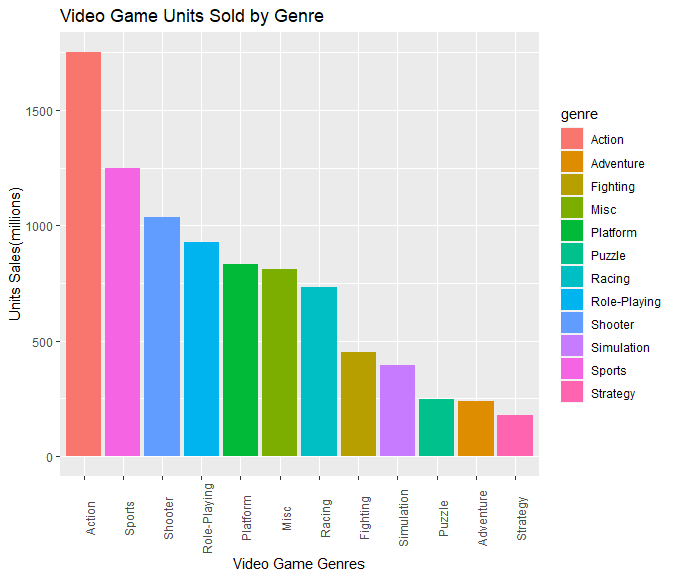
Based on linear modeling outputs, NA Sales again has the greatest influence on the dependent variable of Global Sales. However, we can see that the Japanese sales do not have a statistically significant influence on our global sales with a high p-value of .1468. Furthermore, our Adjusted R Square only shows a significance of 11.82%. Japan also has outliers in their genres, meaning, that in order to enter this market successfully one would have to ensure that it is tailored to Japan’s genre of choice.

**Outcomes and Interpretation**

According to a report published by IBISWorld, the publishing software industry is currently in the growth stage of the life cycle (Cook, 2019). Certain markers indicate this cycle for what it is, including: market expansions, increase in the number of industry operators, and the exorbitant rate of technological change (Cook, 2019). Cook further states that the growth is steady with an expected positive outcome (2019). Meaning, while growth will continue to be exponential, it will also continue a steady uphill climb.

After analysis and attempting to determine what would be our intended target for market entry based on publisher and genre, we looked to find commonalities among the data. Fortunately, Nintendo has made this easy for us. Nintendo was the top grossing publisher in all markets: Global, North America, Japan, and European. This is not surprising, according to WePC, 41.7% of Nintendo’s revenue in 2017 came from North America, while 26.5% and 26.4% were allocated to the Japan and European markets respectively (2020).

Our outcomes suggest that the “Action” genre has the greatest relevance among the top markets, which North America’s sales far exceeding all other markets within this genre type. With a max-sales of 877.3 million allocated to the Action genre. However, with Japan as the outlier, producing a role-playing game would be best to generate the highest profit for the coming year in that market. The focus on the North America and Japanese markets and development of an Action and Role-playing games in those areas will help to make 2020 a successful year for Nintendo. In the European Market, we recommend that Nintendo continue to focus on action games to appeal to the consumers within the market. It should not necessarily be geared towards a Platform, as Sony and Microsoft have yet to eclipse Nintendo in the revenue across all markets. It is also useful to see how the Genre popularity has changed over time, and how that could also affect market entry.



filter(Global\_Sales >= 8 & Global\_Sales <= 40.5) %>%

ggplot()

+

aes(x = Year, fill = Genre, weight = Global\_Sales)

+

geom\_bar()

+

scale\_fill\_hue()

+

labs(x = "Year", y = "Global Sales", title = "Genre Popularity", subtitle = "1984 - 2015")

+

theme\_minimal()

data <- data.frame(market,genre1,Unit)

marketGenre<-ggplot(data, aes(fill=genre1,y=Unit, x=market,col= 'black'))

+

geom\_bar(position="dodge", stat="identity")

+

scale\_color\_brewer(type = 'qual')

Figure 10: Genre Sales in Global Market

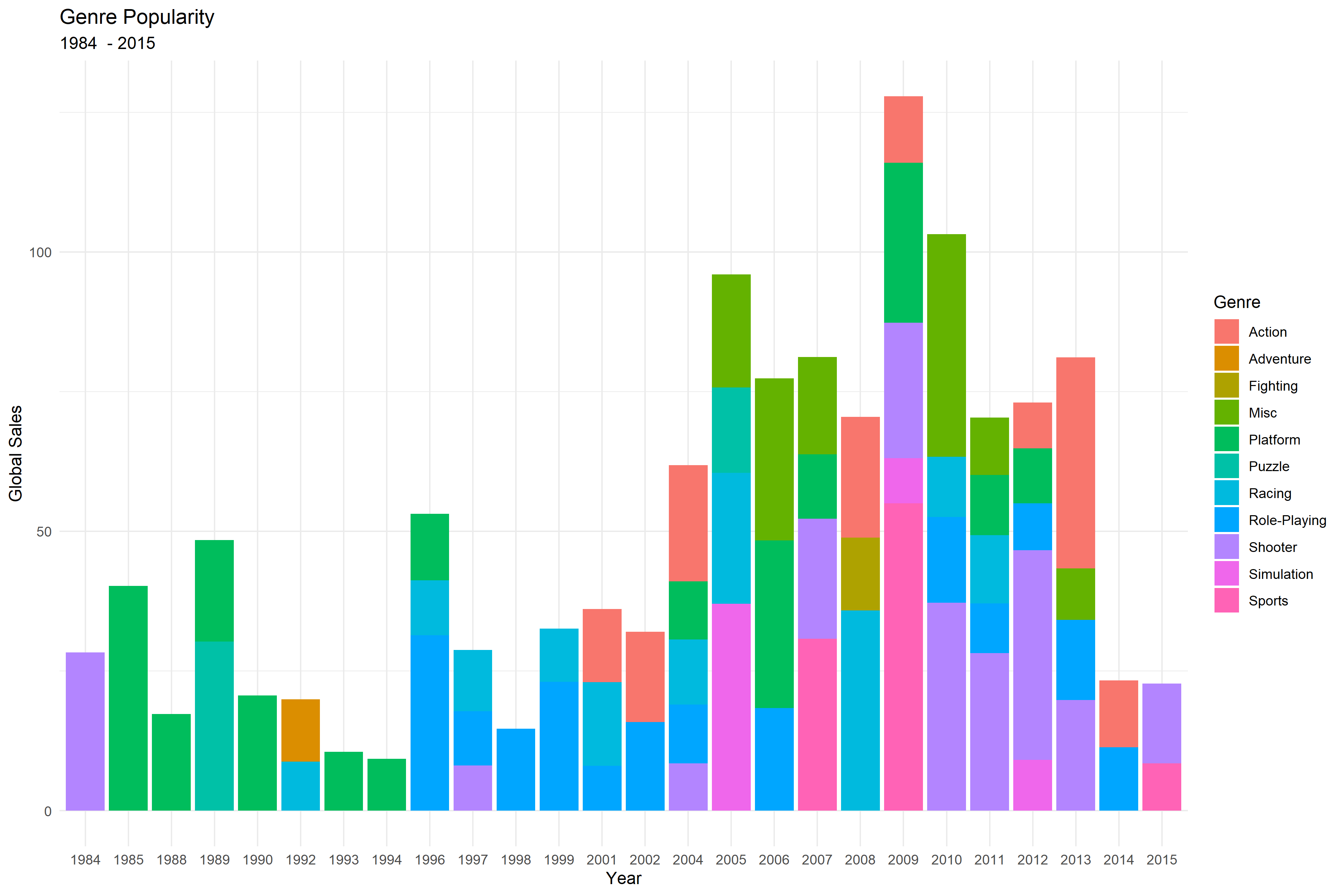


Figure 11: Genre Sales Over Time

**Conclusion**

While the video game industry remains relatively young, it is only expected to further grow in the coming years. Certain aspects of the market can be volatile and are also influenced by the overall economic fluctuations (Cook, 2019). However, with Nintendo focusing on the Action and Role-Playing genres in North America and Japan respectively, we can conclude that their sales will continue to set the precedent across the globe. Furthermore, with a general stepping back, and focusing on the Action genre in Europe, Nintendo should continue to do well in that market. As video games continue to become more accessible to consumers, the sales will continue to rise. Given the outcome, it appears as if Nintendo will continue to remain a dominant force within markets globally for years to come.

**References**

Cifaldi, F. (2012). *Happy 40th birthday, video games.* Retrieved from: <https://web.archive.org/web/20170627094159/http://www.gamasutra.com/view/news/176677/Happy_40th_birthday_video_games.php>, May 29, 2020.

Cook, D. (2019). *Video game software publishing in the U.S. Level up:* *Video game sales through new platforms will boost industry revenue* (Industry Report 51121E)*.* IBISWorld.

Noor, H. (2020). *Video games sales: Video game sales from vgchartz and corresponding ratings from metacritic* (Kaggle, Version 2) [Data set]. Kaggle. https://www.kaggle.com/mohalim/video-games-sales.

Salts, J.S., & Stanton, J.M. (2018). *An introduction to data science.* Sage Publications.

WePC (2020). *2020 video game industry statistics, trends, &data.* Retrieved from: <https://www.wepc.com/news/video-game-statistics/>, May 28, 2020.

Willaert, K. (2018). *In search of the first video game commercial.* Retrieved from: [https://web.archive.org/web/20180112182723/https://gamehistory.org/first-video-game-commercial/](https://web.archive.org/web/20180112182723/https:/gamehistory.org/first-video-game-commercial/), May 30, 2020.

**Appendix A**

#R Code – Unexecuted

VG<-read.csv("~/R/vgsalesV3.csv", header = TRUE)

str(VG)

#removing wii sports because of skewness

VG<-VG[-1,]

VG

str(VG)

tapply(VG$Publisher, VG$Genre, )

rank(VG$Publisher)

str(VG)

#Renumbering Rank

VG$Rank<-c(1:16597)

#Updating Row Names

rownames(VG)<-NULL

View(VG)

#Data Frame for Global Sales/Renaming rows/Renaming columns

Pub1<-tapply(VG$Global\_Sales,VG$Publisher,sum)

Pub2 <- cbind(rownames(Pub1), data.frame(Pub1, row.names=NULL))

names(Pub2)[names(Pub2)=="rownames(Pub1)"]<-"Publisher"

names(Pub2)[names(Pub2)=="Pub1"]<-"Global Sales"

View(Pub2)

#Data Frame for NA Sales/renaming rows/renaming columns

Pub3<-tapply(VG$NA\_Sales,VG$Publisher,sum)

Pub3<-cbind(rownames(Pub3),data.frame(Pub3,row.names=NULL))

names(Pub3)[names(Pub3)=="rownames(Pub3)"]<-"Publisher"

names(Pub3)[names(Pub3)=="Pub3"]<-"North America Sales"

#Data Frame for EU Sales and Renaming/Removing Columns

Pub4<-tapply(VG$EU\_Sales,VG$Publisher,sum)

Pub4<-cbind(rownames(Pub4),data.frame(Pub4,row.names=NULL))

names(Pub4)[names(Pub4)=="rownames(Pub4)"]<-"Publisher"

names(Pub4)[names(Pub4)=="Pub4"]<-"Europe Sales"

#Data Frame for JP Sales/renaming rows/renaming columns

Pub5<-tapply(VG$JP\_Sales,VG$Publisher,sum)

Pub5<-cbind(rownames(Pub5),data.frame(Pub5,row.names=NULL))

names(Pub5)[names(Pub5)=="rownames(Pub5)"]<-"Publisher"

names(Pub5)[names(Pub5)=="Pub5"]<-"Japan Sales"

#Sales by Publisher All-Up Data Frame Merger

pubmerge<-data.frame(c(Pub2,Pub3,Pub4,Pub5))

pubmerge<-pubmerge[,-c(3)]

pubmerge<-pubmerge[,-c(4)]

pubmerge<-pubmerge[,-c(5)]

#Ascending publishers by global sales

pubmerge2<-pubmerge[order(-pubmerge$Global.Sales,pubmerge$Publisher),]

rownames(pubmerge2)<-NULL

pubmerge2[1:3,1]

pm2<-pubmerge2[1:3,1]

#Ascending Publishers by USA Sales

pubmerge3<-pubmerge[order(-pubmerge$North.America.Sales,pubmerge$Publisher),]

rownames(pubmerge3)<-NULL

pubmerge3[1:3,1]

pm3<-pubmerge3[1:3,1]

#Ascending PUblishers by European Sales

pubmerge4<-pubmerge[order(-pubmerge$Europe.Sales,pubmerge$Publisher),]

rownames(pubmerge4)<-NULL

pubmerge4[1:3,1]

pubmerge4[1:3]

pm4<-pubmerge4[1:3,1]

#Ascending Publishers by Japan Sales

pubmerge5<-pubmerge[order(-pubmerge$Japan.Sales,pubmerge$Publisher),]

rownames(pubmerge5)<-NULL

pubmerge5[1:3,1]

pm5<-pubmerge5[1:3,1]

pm6<-data.frame(c(pm2, pm3, pm4, pm5))

#Simplifying publishers

pubmerge2.2<-head(pubmerge2)

pubmerge2.2<-pubmerge2.2[,-c(3:5)]

colnames(pubmerge2.2)

pubmerge3.3<-head(pubmerge3)

colnames(pubmerge3.3)

pubmerge3.3<-pubmerge3.3[,-c(2,4:5)]

colnames(pubmerge3.3)

pubmerge4.4<-head(pubmerge4)

colnames(pubmerge4.4)

pubmerge4.4<-pubmerge4.4[,-c(2:3,5)]

colnames(pubmerge4.4)

pubmerge5.5<-head(pubmerge5)

pubmerge5.5<-pubmerge5.5[,-c(2:4)]

colnames(pubmerge5.5)

puballup<-data.frame(pubmerge2.2,pubmerge3.3,pubmerge4.4,pubmerge5.5)

View(puballup)

pubmelt<-melt(puballup,"Publisher")

View(pubmelt)

pubmelt2<-pubmelt[-c(7:12),]

View(pubmelt2)

rownames(pubmelt2)<-NULL

pubmelt2<-pubmelt2[-c(13:18),]

rownames(pubmelt2)<-NULL

pubmelt2<-pubmelt2[-c(19:24),]

names(pubmelt2)[names(pubmelt2)=="variable"]<-"Region"

names(pubmelt2)[names(pubmelt2)=="value"]<-"Sales"

colnames(pubmelt2)

pubmelt3<-(pubmelt2)

pubmelt3$Region<-gsub("\\."," ",pubmelt3$Region)

pubmelt3$Sales<-as.numeric(pubmelt3$Sales)

str(pubmelt3)

saleshtmap<-ggplot(pubmelt3,aes(x=Region,y=Publisher))+geom\_tile(aes(fill=Sales))+scale\_fill\_gradient(low="blue",high="red")

saleshtmap

salesscatter<-ggplot(pubmelt3)+geom\_point(aes(x=Region,y=Publisher,size=Sales))+scale\_fill\_gradient(low="black",high="orchid")

salesscatter

sd(puballup$Global.Sales)

sd(puballup$North.America.Sales)

sd(puballup$Europe.Sales)

sd(puballup$Japan.Sales)

summary(puballup$Global.Sales)

summary(puballup$North.America.Sales)

summary(puballup$Europe.Sales)

summary(puballup$Japan.Sales)

#make a simplified table based on sales by genre

globalSales<-tapply(VG$Global\_Sales,VG$Genre,sum)

genre<-rownames(globalSales)

GenreSalesGlobal<-data.frame(genre,globalSales)

naSales<-tapply(VG$NA\_Sales,VG$Genre,sum)

euSales<-tapply(VG$EU\_Sales,VG$Genre,sum)

jpSales<-tapply(VG$JP\_Sales,VG$Genre,sum)

otherSales<-tapply(VG$Other\_Sales,VG$Genre,sum)

genreSalesTotal<-data.frame(naSales,euSales,jpSales,otherSales)

#making a fucntion to remove rownames and make factors into strings for the dataframe

RNremove<-function(x){

rownames(x)<-NULL

return(x)

}

RNremove(VG)

RNremove(GenreSalesGlobal)

#converting genre column from factor to character

i <- sapply(GenreSalesGlobal, is.factor)

GenreSalesGlobal[i] <- lapply(GenreSalesGlobal[i], as.character)

str(GenreSalesGlobal)

#analysis of genre sales by market

statsVG<-function(x){

a<-mean(x)

b<-median(x)

c<-min(x)

d<-max(x)

e<-sd(x)

f<-quantile(x,probs = 0.05)

g<-quantile(x,probs = 0.95)

cat("mean:" ,a, "median:" ,b, "min:" ,c, "max" ,d, "sd:",e,"quantile(0.05-0.95):",f,"--",g)

}

statsVG(genreSalesTotal$naSales)

statsVG(genreSalesTotal$euSales)

statsVG(genreSalesTotal$jpSales)

statsVG(genreSalesTotal$otherSales)

#creating a bar graph to compare genre success based on region

gGlobal<-ggplot(GenreSalesGlobal, aes(x = reorder(genre, -globalSales),y=globalSales, fill = genre)) + geom\_bar(stat = "identity")

gGlobal<-gGlobal + labs(y= 'Units Sales(millions)', x ='Video Game Genres') + ggtitle('Video Game Units Sold by Genre')

gGlobal<-gGlobal + theme(axis.text.x = element\_text(angle = 90, hjust = 0.5))

gGlobal

#create a bar graph to show sales by genre across markets

#make row names into a column called genre

genreSalesTotal$genre<-rownames(genreSalesTotal)

rownames(genreSalesTotal)<-NULL

# compare genre performance across markets

market<- c(rep("NaSales" , 12) , rep("EUSales" , 12) , rep("JPSales" , 12) , rep("OtherSales" , 12) )

genre1 <- rep(c("Action" , "Adventure","Fighting", "Misc", "Platform","Puzzle","Racing","Role-Playing",

"Shooter", "Simulation","sports", "Strategy") , 4)

Unit <- c(877.83,105.80,223.59,410.24,447.05,123.78,359.42,327.28,582.60,183.31,683.35,68.70,525.00,

64.13,101.32,215.98,201.63,50.78,238.39,188.06,313.27,113.38,376.85,45.34,159.95,52.07,

87.35,107.76,130.77,57.31,56.69,352.31,38.28,63.70,135.37,49.46,187.38,16.81,36.68,75.32,

51.59,12.55,77.27,59.61,102.69,31.52,134.97,11.36)

data <- data.frame(market,genre1,Unit)

marketGenre<-ggplot(data, aes(fill=genre1,y=Unit, x=market,col= 'black')) +

geom\_bar(position="dodge", stat="identity") + scale\_color\_brewer(type = 'qual')

marketGenre

#TOP 3 GENRES:

#EU: 1.Action, 2.Sports, and 3.Shooters

#JP: 1.Role-Playing, 2.Action, and 3.Sports

#NA: 1.Action, 2.Sports, and 3.Shooters

#Other:

#compare sales over time

#Create matrix of total sales in each market by year. NA'S included

NAsalesByDate<-tapply(VG$NA\_Sales,list(VG$Year,VG$Genre),sum)

JPsalesByDate<-tapply(VG$JP\_Sales,list(VG$Year,VG$Genre),sum)

EUsalesByDate<-tapply(VG$EU\_Sales,list(VG$Year,VG$Genre),sum)

OthersalesByDate<-tapply(VG$Other\_Sales,list(VG$Year,VG$Genre),sum)

#change NA'S to zeros

NAsalesByDate[is.na(NAsalesByDate)]<-0

JPsalesByDate[is.na(JPsalesByDate)]<-0

EUsalesByDate[is.na(EUsalesByDate)]<-0

OthersalesByDate[is.na(NAsalesByDate)]<-0

#parce global data

VG<-na.omit(VG)

gGenreYear<-sqldf("SELECT sum(Global\_Sales), ")

g80<-sqldf("SELECT sum(Global\_Sales),Year FROM VG WHERE Year<1990 group by Year")

g80

g90<-sqldf("SELECT sum(Global\_Sales),Year FROM VG WHERE Year<2000 AND Year>1989 group by Year")

g90

g00<-sqldf("SELECT sum(Global\_Sales),Year FROM VG WHERE Year<2010 AND Year>1999 group by Year")

g00

g10<-sqldf("SELECT sum(Global\_Sales),Year FROM VG WHERE Year<2021 AND Year>2009 group by Year")

g10

gTotal<-sqldf("SELECT sum(Global\_Sales), Year FROM VG group by Year")

gTotal

nintendoGlobal<-sqldf("SELECT sum(Global\_Sales),Publisher,Year FROM VG WHERE Publisher = 'Nintendo' group by Year")

nintendoGlobal

nintendoNA<-sqldf("SELECT sum(NA\_Sales),Publisher,Year FROM VG WHERE Publisher = 'Nintendo' group by Year")

nintendoNA

nintendoJP<-sqldf("SELECT sum(JP\_Sales),Publisher,Year FROM VG WHERE Publisher = 'Nintendo' group by Year")

nintendoJP

nintendoEU<-sqldf("SELECT sum(EU\_Sales),Publisher,Year FROM VG WHERE Publisher = 'Nintendo' group by Year")

nintendoEU

nintendoOther<-sqldf("SELECT sum(Other\_Sales),Publisher,Year FROM VG WHERE Publisher = 'Nintendo' group by Year")

nintendoOther

nintendoTotal<-as.data.frame(nintendoNA,nintendoEU,nintendoJP,nintendoOther)

#rename axis lables

colnames(gTotal)<-c("Global\_Sales","Year")

colnames(g80)<-c("Global\_Sales","Year")

colnames(g90)<-c("Global\_Sales","Year")

colnames(g00)<-c("Global\_Sales","Year")

colnames(g10)<-c("Global\_Sales","Year")

#plot the time lines of global sales

g80P<-ggplot(g80,aes(y=Global\_Sales, x=Year, group = 1))+geom\_line()+geom\_point(data = g80)

g80P<-g80P + ggtitle("Global Sales in the 1980's")

g80P

g90P<-ggplot(g90,aes(y=Global\_Sales, x=Year, group = 1))+geom\_line(color="orchid")+geom\_point(color="blue",data = g90)

g90P<-g90P + ggtitle("Global Sales in the 1990's") + theme(plot.title=element\_text(hjust=.5))+ylab("Global Sales in Millions")

g90P

g00P<-ggplot(g00,aes(y=Global\_Sales, x=Year, group = 1))+geom\_line(color="orchid")+geom\_point(color="blue",data = g00)

g00P<-g00P + ggtitle("Global Sales in the 2000's") + theme(plot.title=element\_text(hjust=.5))+ylab("Global Sales in Millions")

g00P

g10P<-ggplot(g10,aes(y=Global\_Sales, x=Year, group = 1))+geom\_line(color="orchid")+geom\_point(color="blue",data = g10)

g10P<-g10P + ggtitle("Global Sales in the 2010's") + theme(plot.title=element\_text(hjust=.5))+ylab("Global Sales in Millions")

g10PgTotalP<-ggplot(gTotal,aes(y=Global\_Sales, x=Year, group = 1))+geom\_line(color="orchid")+geom\_point(color="blue",data = gTotal)

gTotalP<-gTotalP + ggtitle("Global Sales from 1980 to 2020") + theme(plot.title=element\_text(hjust=.5))+ylab("Global Sales in Millions")

gTotalP

#mike's code#

devtools::install\_github("robj")

#### Call esquisse ggplot plug-in ######

esquisse::esquisser()

###Global Sales by Genre – Fill Platform & Color Platform

ggplot(VG) +

aes(x = Genre, fill = Platform, colour = Platform, weight = Global\_Sales) +

geom\_bar() +

scale\_fill\_hue() +

scale\_color\_hue() +

theme\_minimal()

###Global Sales by Publisher – Fill Platform & faceted by year \* (slice by decade?)

VG %>%

filter(Global\_Sales >= 9 & Global\_Sales <= 40.5) %>%

ggplot() +

aes(x = Publisher, fill = Platform, weight = Global\_Sales) +

geom\_bar() +

scale\_fill\_hue() +

labs(y = "Global Sales") +

coord\_flip() +

theme\_minimal() +

facet\_wrap(vars(Year))

#Global Sales Platform V Publisher

VG %>%

filter(Global\_Sales >= 8 & Global\_Sales <= 40.5) %>%

ggplot() +

aes(x = Platform, fill = Publisher, weight = Global\_Sales) +

geom\_bar() +

scale\_fill\_hue() +

labs(y = "Global Sales", title = "Global Sales ", subtitle = "Plaform V Publisher") +

theme\_minimal()

#Genre Popularity – Sales over Time

VG %>%

filter(Global\_Sales >= 8 & Global\_Sales <= 40.5) %>%

ggplot() +

aes(x = Year, fill = Genre, weight = Global\_Sales) +

geom\_bar() +

scale\_fill\_hue() +

labs(x = "Year", y = "Global Sales", title = "Genre Popularity", subtitle = "1984 - 2015") +

theme\_minimal()

#Platform Popularity -

VG %>%

filter(Global\_Sales >= 8 & Global\_Sales <= 40.5) %>%

ggplot() +

aes(x = Year, fill = Platform, weight = Global\_Sales) +

geom\_bar() +

scale\_fill\_hue() +

labs(x = "Year", y = "Global Sales", title = "Platform Popularity", subtitle = "1984 - 2015") +

theme\_minimal()

**Appendix B**

R Code – Executed

77 Atlus 12.80 4.20

78 Titus 3.31 1.96

79 Kalypso Media 2.40 1.00

80 Oxygen Interactive 3.41 2.01

81 Video System 2.17 0.67

82 Ocean 3.59 2.06

83 Telltale Games 2.70 1.28

84 RTL 2.01 0.83

85 Midas Interactive Entertainment 2.72 1.34

86 Hudson Soft 23.02 2.55

87 PQube 4.28 1.91

88 Slightly Mad Studios 1.37 0.25

89 SouthPeak Games 6.31 4.77

90 Activision Value 6.98 5.53

91 Arena Entertainment 4.72 3.70

92 Xplosiv 1.64 0.61

93 Red Storm Entertainment 2.32 1.43

94 Milestone S.r.l. 1.27 0.23

95 ASC Games 2.02 1.12

96 Tomy Corporation 4.96 1.84

97 Hello Games 1.60 0.58

98 Natsume 3.70 2.52

99 Ubisoft Annecy 4.48 1.27

100 Crystal Dynamics 1.72 0.95

101 System 3 Arcade Software 1.79 0.96

102 Valve 1.73 0.83

103 Bigben Interactive 0.85 0.10

104 Agetec 1.93 0.78

105 Paradox Interactive 0.70 0.05

106 O-Games 1.10 0.44

107 Avanquest Software 1.68 0.95

108 LEGO Media 1.56 1.01

109 Mastertronic 1.70 1.07

110 Touchstone 1.59 0.88

111 Home Entertainment Suppliers 0.70 0.09

112 Kemco 2.19 1.61

113 Success 1.92 1.12

114 JoWood Productions 1.50 0.87

115 Gathering of Developers 1.20 0.61

116 Metro 3D 1.81 0.77

117 Rage Software 1.41 0.81

118 Alchemist 2.62 0.59

119 Black Label Games 1.20 0.59

120 Ghostlight 2.79 1.22

121 UEP Systems 2.26 1.52

122 ASCII Entertainment 10.90 0.70

123 BMG Interactive Entertainment 1.18 0.67

124 Palcom 4.17 3.38

125 Asylum Entertainment 0.58 0.09

126 Jaleco 1.95 0.79

127 NewKidCo 1.31 0.89

128 Gremlin Interactive Ltd 1.08 0.66

129 Mattel Interactive 3.47 3.05

130 Activision Blizzard 0.48 0.08

131 Genki 1.47 0.49

132 Accolade 1.66 1.25

133 CTO SpA 0.90 0.50

134 PopCap Games 1.56 1.10

135 JVC 2.78 2.25

136 P2 Games 0.39 0.02

137 Spike 6.14 3.32

138 TalonSoft 0.88 0.49

139 Astragon 0.62 0.24

140 Enix Corporation 33.74 0.76

141 Kadokawa Shoten 3.77 0.46

142 Sammy Corporation 2.73 0.41

143 Foreign Media Games 0.62 0.27

144 PlayV 0.45 0.10

145 Square EA 0.80 0.45

146 Trion Worlds 1.13 0.72

147 Jester Interactive 1.13 0.70

148 Square 3.65 0.53

149 Illusion Softworks 0.70 0.34

150 Imagic 4.82 4.51

151 Level 5 12.16 0.20

152 Parker Bros. 4.97 4.65

153 Mud Duck Productions 0.83 0.52

154 Sunsoft 0.96 0.50

155 Tru Blu Entertainment 0.28 0.01

156 NovaLogic 0.79 0.48

157 Daedalic 0.30 0.00

158 Excalibur Publishing 0.28 0.00

159 Wargaming.net 0.29 0.00

160 Marvelous Entertainment 2.34 0.80

161 Brash Entertainment 1.21 0.89

162 From Software 1.65 0.35

163 Wanadoo 0.64 0.30

164 Banpresto 17.41 0.41

165 bitComposer Games 0.44 0.16

166 Game Factory 3.74 3.31

167 CCP 0.21 0.00

168 Starfish 0.71 0.39

169 Alternative Software 0.21 0.00

170 Nobilis 1.95 1.61

171 Coleco 3.07 2.87

172 Marvelous Interactive 4.18 1.61

173 Merscom LLC 0.20 0.00

174 XS Games 1.46 1.18

175 Electronic Arts Victor 0.43 0.23

176 Rebellion Developments 0.29 0.08

177 Knowledge Adventure 1.46 1.23

178 Pioneer LDC 0.75 0.23

179 Playlogic Game Factory 1.21 0.94

180 Swing! Entertainment 0.48 0.28

181 TYO 0.42 0.23

182 989 Sports 0.37 0.21

183 Conspiracy Entertainment 0.81 0.63

184 DreamCatcher Interactive 0.96 0.75

185 Microids 0.38 0.23

186 Nordcurrent 0.58 0.39

187 SNK Playmore 0.96 0.23

188 Big Ben Interactive 0.98 0.78

189 Funbox Media 0.90 0.71

190 Game Life 0.29 0.13

191 Gamebridge 0.73 0.30

192 Magix 0.34 0.17

193 Rebellion 0.49 0.35

194 Reef Entertainment 0.33 0.17

195 Tetris Online 0.42 0.20

196 The Adventure Company 0.52 0.34

197 U.S. Gold 0.34 0.19

198 TDK Core 0.62 0.24

199 Avalon Interactive 0.49 0.34

200 SCS Software 0.14 0.00

201 Time Warner Interactive 0.33 0.18

202 Core Design Ltd. 0.29 0.16

203 Blast! Entertainment Ltd 0.37 0.23

204 Daedalic Entertainment 0.12 0.00

205 Hamster Corporation 0.31 0.16

206 HMH Interactive 0.13 0.00

207 Playmates 0.29 0.16

208 Yacht Club Games 0.31 0.16

209 20th Century Fox Video Games 1.93 1.82

210 Bohemia Interactive 0.12 0.00

211 Enterbrain 2.23 0.12

212 Funcom 0.19 0.07

213 Jack of All Games 0.42 0.31

214 Magical Company 0.40 0.15

215 Slitherine Software 0.21 0.08

216 Storm City Games 1.53 1.35

217 System 3 0.21 0.08

218 Telstar 0.28 0.16

219 TopWare Interactive 0.12 0.00

220 American Softworks 0.24 0.13

221 Ascaron Entertainment GmbH 0.11 0.00

222 Harmonix Music Systems 0.80 0.60

223 Havas Interactive 0.12 0.01

224 Hip Interactive 0.35 0.26

225 Pacific Century Cyber Works 0.26 0.11

226 Taito 2.83 0.41

227 Tigervision 1.43 1.33

228 ValuSoft 0.39 0.26

229 Destination Software, Inc 0.86 0.75

230 Idea Factory International 0.60 0.23

231 Neko Entertainment 0.30 0.20

232 On Demand 0.21 0.12

233 Valcon Games 0.42 0.31

234 Vic Tokai 0.28 0.19

235 Victor Interactive 0.89 0.11

236 Adeline Software 0.18 0.10

237 Indie Games 0.32 0.09

238 Liquid Games 0.26 0.19

239 SSI 0.19 0.11

240 Sting 0.65 0.40

241 Tivola 0.17 0.08

242 Tripwire Interactive 0.13 0.04

243 1C Company 0.10 0.01

244 Revolution Software 0.08 0.00

245 Vatical Entertainment 0.22 0.14

246 AQ Interactive 0.56 0.34

247 Crimson Cow 0.07 0.00

248 7G//AMES 0.06 0.00

249 CDV Software Entertainment 0.18 0.11

250 EA Games 0.15 0.07

251 inXile Entertainment 0.09 0.02

252 Licensed 4U 0.06 0.00

253 Pack In Soft 0.48 0.25

254 Takara 3.44 0.16

255 Telegames 1.03 0.91

256 Tradewest 0.09 0.02

257 Xicat Interactive 0.16 0.09

258 ArtDink 0.75 0.07

259 Destineer 4.37 4.06

260 DreamWorks Interactive 0.13 0.07

261 Focus Multimedia 0.05 0.00

262 FunSoft 0.13 0.07

263 Funsta 0.19 0.13

264 Gust 1.09 0.15

265 Hudson Entertainment 1.69 0.33

266 Iceberg Interactive 0.07 0.02

267 Idea Factory 4.00 0.53

268 iWin 0.06 0.00

269 Mad Catz 0.31 0.24

270 Mastiff 2.23 2.04

271 Media Rings 0.65 0.14

272 Milestone S.r.l 0.06 0.00

273 Mystique 0.82 0.76

274 Quelle 0.87 0.81

275 Zushi Games 1.84 1.67

276 10TACLE Studios 0.11 0.07

277 49Games 0.04 0.00

278 Ackkstudios 0.84 0.13

279 Alawar Entertainment 0.05 0.00

280 Aspyr 0.64 0.55

281 Big Fish Games 0.04 0.00

282 Data Age 0.70 0.66

283 Devolver Digital 0.04 0.00

284 DSI Games 0.34 0.29

285 Evolved Games 0.33 0.26

286 GameMill Entertainment 1.41 1.28

287 Human Entertainment 3.02 0.06

288 Interplay Productions 0.10 0.05

289 Laguna 1.55 0.16

290 LSP Games 0.13 0.08

291 Mamba Games 0.05 0.00

292 Men-A-Vision 0.77 0.72

293 Monte Christo Multimedia 0.05 0.00

294 Myelin Media 0.14 0.09

295 PopTop Software 0.11 0.06

296 responDESIGN 0.14 0.09

297 SPS 0.12 0.07

298 Syscom 0.27 0.06

299 Team17 Software 0.05 0.00

300 Universal Gamex 0.63 0.58

301 Xseed Games 0.89 0.41

302 2D Boy 0.04 0.00

303 Answer Software 0.50 0.46

304 Arika 0.10 0.04

305 Blue Byte 0.03 0.00

306 Compile Heart 1.37 0.50

307 CPG Products 0.54 0.50

308 Culture Publishers 0.09 0.05

309 Easy Interactive 0.03 0.00

310 Elite 0.07 0.04

311 Encore 0.07 0.04

312 Headup Games 0.03 0.00

313 Imagineer 2.65 0.09

314 Just Flight 0.04 0.00

315 MC2 Entertainment 0.06 0.03

316 Origin Systems 0.09 0.05

317 Scholastic Inc. 2.88 2.67

318 SNK 2.59 0.04

319 Takara Tomy 2.72 0.76

320 Ultravision 0.47 0.44

321 Wizard Video Games 0.62 0.58

322 Xing Entertainment 0.25 0.04

323 Aerosoft 0.03 0.01

324 Agatsuma Entertainment 0.31 0.22

325 Altron 0.07 0.05

326 CBS Electronics 0.31 0.29

327 Codemasters Online 0.02 0.00

328 ESP 0.59 0.03

329 FuRyu 1.31 0.11

330 Gamecock 0.18 0.16

331 GameTek 0.10 0.08

332 GOA 0.02 0.00

333 Graphsim Entertainment 0.02 0.00

[ reached 'max' / getOption("max.print") -- omitted 246 rows ]

> pm4<-pubmerge4[1:3,1]

>

> #Ascending Publishers by Japan Sales

> pubmerge5<-pubmerge[order(-pubmerge$Japan.Sales,pubmerge$Publisher),]

> rownames(pubmerge5)<-NULL

> pubmerge5[1:3,1]

[1] Nintendo Namco Bandai Games

[3] Konami Digital Entertainment

579 Levels: 10TACLE Studios 1C Company 20th Century Fox Video Games 2D Boy ... Zushi Games

> pm5<-pubmerge5[1:3,1]

>

>

> pm6<-data.frame(c(pm2, pm3, pm4, pm5))

>

> #Simplifying publishers

> pubmerge2.2<-head(pubmerge2)

> pubmerge2.2<-pubmerge2.2[,-c(3:5)]

> colnames(pubmerge2.2)

[1] "Publisher" "Global.Sales"

>

> pubmerge3.3<-head(pubmerge3)

> colnames(pubmerge3.3)

[1] "Publisher" "Global.Sales" "North.America.Sales"

[4] "Europe.Sales" "Japan.Sales"

> pubmerge3.3<-pubmerge3.3[,-c(2,4:5)]

> colnames(pubmerge3.3)

[1] "Publisher" "North.America.Sales"

>

> pubmerge4.4<-head(pubmerge4)

> colnames(pubmerge4.4)

[1] "Publisher" "Global.Sales" "North.America.Sales"

[4] "Europe.Sales" "Japan.Sales"

> pubmerge4.4<-pubmerge4.4[,-c(2:3,5)]

> colnames(pubmerge4.4)

[1] "Publisher" "Europe.Sales"

>

>

> pubmerge5.5<-head(pubmerge5)

> pubmerge5.5<-pubmerge5.5[,-c(2:4)]

> colnames(pubmerge5.5)

[1] "Publisher" "Japan.Sales"

>

>

>

> puballup<-data.frame(pubmerge2.2,pubmerge3.3,pubmerge4.4,pubmerge5.5)

> View(puballup)

> pubmelt<-melt(puballup,"Publisher")

Warning messages:

1: In melt(puballup, "Publisher") :

The melt generic in data.table has been passed a data.frame and will attempt to redirect to the relevant reshape2 method; please note that reshape2 is deprecated, and this redirection is now deprecated as well. To continue using melt methods from reshape2 while both libraries are attached, e.g. melt.list, you can prepend the namespace like reshape2::melt(puballup). In the next version, this warning will become an error.

2: attributes are not identical across measure variables; they will be dropped

> View(pubmelt)

> pubmelt2<-pubmelt[-c(7:12),]

> View(pubmelt2)

>

> rownames(pubmelt2)<-NULL

> pubmelt2<-pubmelt2[-c(13:18),]

> rownames(pubmelt2)<-NULL

> pubmelt2<-pubmelt2[-c(19:24),]

>

>

> names(pubmelt2)[names(pubmelt2)=="variable"]<-"Region"

> names(pubmelt2)[names(pubmelt2)=="value"]<-"Sales"

> colnames(pubmelt2)

[1] "Publisher" "Region" "Sales"

>

>

> pubmelt3<-(pubmelt2)

> pubmelt3$Region<-gsub("\\."," ",pubmelt3$Region)

> pubmelt3$Sales<-as.numeric(pubmelt3$Sales)

> str(pubmelt3)

'data.frame': 24 obs. of 3 variables:

$ Publisher: Factor w/ 579 levels "10TACLE Studios",..: 369 139 17 465 533 499 369 139 17 465 ...

$ Region : chr "Global Sales" "Global Sales" "Global Sales" "Global Sales" ...

$ Sales : num 1704 1110 727 607 475 ...

>

> saleshtmap<-ggplot(pubmelt3,aes(x=Region,y=Publisher))+geom\_tile(aes(fill=Sales))+scale\_fill\_gradient(low="blue",high="red")

> saleshtmap

>

> salesscatter<-ggplot(pubmelt3)+geom\_point(aes(x=Region,y=Publisher,size=Sales))+scale\_fill\_gradient(low="black",high="orchid")

> salesscatter

>

>

> sd(puballup$Global.Sales)

[1] 492.5489

> sd(puballup$North.America.Sales)

[1] 222.9274

> sd(puballup$Europe.Sales)

[1] 112.8753

> sd(puballup$Japan.Sales)

[1] 152.2639

>

> summary(puballup$Global.Sales)

Min. 1st Qu. Median Mean 3rd Qu. Max.

399.7 507.8 667.3 837.1 1014.4 1703.6

> summary(puballup$North.America.Sales)

Min. 1st Qu. Median Mean 3rd Qu. Max.

220.5 256.4 347.5 423.2 553.7 775.4

> summary(puballup$Europe.Sales)

Min. 1st Qu. Median Mean 3rd Qu. Max.

118.1 169.4 201.6 240.9 332.3 389.7

> summary(puballup$Japan.Sales)

Min. 1st Qu. Median Mean 3rd Qu. Max.

57.03 69.58 82.70 144.87 118.13 451.65

> #make a simplified table based on sales by genre

> globalSales<-tapply(VG$Global\_Sales,VG$Genre,sum)

> genre<-rownames(globalSales)

> GenreSalesGlobal<-data.frame(genre,globalSales)

> naSales<-tapply(VG$NA\_Sales,VG$Genre,sum)

> euSales<-tapply(VG$EU\_Sales,VG$Genre,sum)

> jpSales<-tapply(VG$JP\_Sales,VG$Genre,sum)

> otherSales<-tapply(VG$Other\_Sales,VG$Genre,sum)

> genreSalesTotal<-data.frame(naSales,euSales,jpSales,otherSales)

> #making a fucntion to remove rownames and make factors into strings for the dataframe

> RNremove<-function(x){

+ rownames(x)<-NULL

+ return(x)

+ }

> RNremove(VG)

Rank Name Platform Year Genre

1 1 Super Mario Bros. NES 1985 Platform

2 2 Mario Kart Wii Wii 2008 Racing

3 3 Wii Sports Resort Wii 2009 Sports

4 4 Pokémon Red/Pokémon Blue GB 1996 Role-Playing

5 5 Tetris GB 1989 Puzzle

6 6 New Super Mario Bros. DS 2006 Platform

7 7 Wii Play Wii 2006 Misc

8 8 New Super Mario Bros. Wii Wii 2009 Platform

9 9 Duck Hunt NES 1984 Shooter

10 10 Nintendogs DS 2005 Simulation

11 11 Mario Kart DS DS 2005 Racing

12 12 Pokémon Gold/Pokémon Silver GB 1999 Role-Playing

13 13 Wii Fit Wii 2007 Sports

14 14 Wii Fit Plus Wii 2009 Sports

15 15 Kinect Adventures! X360 2010 Misc

16 16 Grand Theft Auto V PS3 2013 Action

17 17 Grand Theft Auto: San Andreas PS2 2004 Action

18 18 Super Mario World SNES 1990 Platform

19 19 Brain Age: Train Your Brain in Minutes a Day DS 2005 Misc

20 20 Pokémon Diamond/Pokémon Pearl DS 2006 Role-Playing

21 21 Super Mario Land GB 1989 Platform

22 22 Super Mario Bros. 3 NES 1988 Platform

23 23 Grand Theft Auto V X360 2013 Action

24 24 Grand Theft Auto: Vice City PS2 2002 Action

25 25 Pokémon Ruby/Pokémon Sapphire GBA 2002 Role-Playing

26 26 Pokémon Black/Pokémon White DS 2010 Role-Playing

27 27 Brain Age 2: More Training in Minutes a Day DS 2005 Puzzle

28 28 Gran Turismo 3: A-Spec PS2 2001 Racing

29 29 Call of Duty: Modern Warfare 3 X360 2011 Shooter

30 30 Pokémon Yellow: Special Pikachu Edition GB 1998 Role-Playing

31 31 Call of Duty: Black Ops X360 2010 Shooter

32 32 Pokémon X/Pokémon Y 3DS 2013 Role-Playing

33 33 Call of Duty: Black Ops 3 PS4 2015 Shooter

34 34 Call of Duty: Black Ops II PS3 2012 Shooter

35 35 Call of Duty: Black Ops II X360 2012 Shooter

36 36 Call of Duty: Modern Warfare 2 X360 2009 Shooter

37 37 Call of Duty: Modern Warfare 3 PS3 2011 Shooter

38 38 Grand Theft Auto III PS2 2001 Action

39 39 Super Smash Bros. Brawl Wii 2008 Fighting

40 40 Call of Duty: Black Ops PS3 2010 Shooter

41 41 Animal Crossing: Wild World DS 2005 Simulation

42 42 Mario Kart 7 3DS 2011 Racing

43 43 Halo 3 X360 2007 Shooter

44 44 Grand Theft Auto V PS4 2014 Action

45 45 Pokémon HeartGold/Pokémon SoulSilver DS 2009 Action

46 46 Super Mario 64 N64 1996 Platform

47 47 Gran Turismo 4 PS2 2004 Racing

48 48 Super Mario Galaxy Wii 2007 Platform

49 49 Pokémon Omega Ruby/Pokémon Alpha Sapphire 3DS 2014 Role-Playing

50 50 Super Mario Land 2: 6 Golden Coins GB 1992 Adventure

51 51 Grand Theft Auto IV X360 2008 Action

52 52 Gran Turismo PS 1997 Racing

53 53 Super Mario 3D Land 3DS 2011 Platform

54 54 Gran Turismo 5 PS3 2010 Racing

55 55 Call of Duty: Modern Warfare 2 PS3 2009 Shooter

56 56 Grand Theft Auto IV PS3 2008 Action

57 57 Super Mario All-Stars SNES 1993 Platform

58 58 Pokémon FireRed/Pokémon LeafGreen GBA 2004 Role-Playing

59 59 Super Mario 64 DS 2004 Platform

60 60 Just Dance 3 Wii 2011 Misc

61 61 Call of Duty: Ghosts X360 2013 Shooter

62 62 Halo: Reach X360 2010 Shooter

63 63 Mario Kart 64 N64 1996 Racing

64 64 New Super Mario Bros. 2 3DS 2012 Platform

65 65 Halo 4 X360 2012 Shooter

66 66 Final Fantasy VII PS 1997 Role-Playing

67 67 Call of Duty: Ghosts PS3 2013 Shooter

68 68 Just Dance 2 Wii 2010 Misc

69 69 Gran Turismo 2 PS 1999 Racing

70 70 Call of Duty 4: Modern Warfare X360 2007 Shooter

71 71 Donkey Kong Country SNES 1994 Platform

72 72 Minecraft X360 2013 Misc

73 73 Animal Crossing: New Leaf 3DS 2012 Simulation

74 74 Mario Party DS DS 2007 Misc

75 75 The Elder Scrolls V: Skyrim X360 2011 Role-Playing

76 76 Super Mario Kart SNES 1992 Racing

77 77 FIFA 16 PS4 2015 Sports

78 78 Wii Party Wii 2010 Misc

79 79 Halo 2 XB 2004 Shooter

80 80 Mario Party 8 Wii 2007 Misc

81 81 Pokémon Black 2/Pokémon White 2 DS 2012 Role-Playing

82 82 FIFA Soccer 13 PS3 2012 Action

83 83 The Sims 3 PC 2009 Simulation

84 84 GoldenEye 007 N64 1997 Shooter

85 85 Mario & Sonic at the Olympic Games Wii 2007 Sports

86 86 Final Fantasy X PS2 2001 Role-Playing

87 87 Final Fantasy VIII PS 1999 Role-Playing

88 88 Pokémon Platinum Version DS 2008 Role-Playing

89 89 Pac-Man 2600 1982 Puzzle

90 90 Grand Theft Auto: Liberty City Stories PSP 2005 Action

Publisher NA\_Sales EU\_Sales JP\_Sales Other\_Sales Global\_Sales

1 Nintendo 29.08 3.58 6.81 0.77 40.24

2 Nintendo 15.85 12.88 3.79 3.31 35.83

3 Nintendo 15.75 11.01 3.28 2.96 33.00

4 Nintendo 11.27 8.89 10.22 1.00 31.38

5 Nintendo 23.20 2.26 4.22 0.58 30.26

6 Nintendo 11.38 9.23 6.50 2.90 30.01

7 Nintendo 14.03 9.20 2.93 2.85 29.01

8 Nintendo 14.59 7.06 4.70 2.26 28.61

9 Nintendo 26.93 0.63 0.28 0.47 28.31

10 Nintendo 9.07 11.00 1.93 2.75 24.75

11 Nintendo 9.81 7.57 4.13 1.92 23.43

12 Nintendo 9.00 6.18 7.20 0.71 23.09

13 Nintendo 8.94 8.03 3.60 2.15 22.72

14 Nintendo 9.09 8.59 2.53 1.79 22.00

15 Microsoft Game Studios 14.97 4.94 0.24 1.67 21.82

16 Take-Two Interactive 7.01 9.27 0.97 4.14 21.39

17 Take-Two Interactive 9.43 0.40 0.41 10.57 20.81

18 Nintendo 12.78 3.75 3.54 0.55 20.62

19 Nintendo 4.75 9.26 4.16 2.05 20.22

20 Nintendo 6.42 4.52 6.04 1.37 18.35

21 Nintendo 10.83 2.71 4.18 0.42 18.14

22 Nintendo 9.54 3.44 3.84 0.46 17.28

23 Take-Two Interactive 9.63 5.31 0.06 1.38 16.38

24 Take-Two Interactive 8.41 5.49 0.47 1.78 16.15

25 Nintendo 6.06 3.90 5.38 0.50 15.84

26 Nintendo 5.57 3.28 5.65 0.82 15.32

27 Nintendo 3.44 5.36 5.32 1.18 15.30

28 Sony Computer Entertainment 6.85 5.09 1.87 1.16 14.97

29 Activision 9.03 4.28 0.13 1.32 14.76

30 Nintendo 5.89 5.04 3.12 0.59 14.64

31 Activision 9.67 3.73 0.11 1.13 14.64

32 Nintendo 5.17 4.05 4.34 0.79 14.35

33 Activision 5.77 5.81 0.35 2.31 14.24

34 Activision 4.99 5.88 0.65 2.52 14.04

35 Activision 8.25 4.30 0.07 1.12 13.74

36 Activision 8.52 3.63 0.08 1.29 13.52

37 Activision 5.54 5.82 0.49 1.62 13.47

38 Take-Two Interactive 6.99 4.51 0.30 1.30 13.10

39 Nintendo 6.75 2.61 2.66 1.02 13.04

40 Activision 5.98 4.44 0.48 1.83 12.73

41 Nintendo 2.55 3.52 5.33 0.88 12.28

42 Nintendo 4.74 3.91 2.67 0.89 12.21

43 Microsoft Game Studios 7.97 2.83 0.13 1.21 12.14

44 Take-Two Interactive 3.80 5.81 0.36 2.02 11.99

45 Nintendo 4.40 2.77 3.96 0.77 11.90

46 Nintendo 6.91 2.85 1.91 0.23 11.90

47 Sony Computer Entertainment 3.01 0.01 1.10 7.53 11.65

48 Nintendo 6.16 3.40 1.20 0.76 11.52

49 Nintendo 4.23 3.37 3.08 0.65 11.33

50 Nintendo 6.16 2.04 2.69 0.29 11.18

51 Take-Two Interactive 6.76 3.10 0.14 1.03 11.03

52 Sony Computer Entertainment 4.02 3.87 2.54 0.52 10.95

53 Nintendo 4.89 2.99 2.13 0.78 10.79

54 Sony Computer Entertainment 2.96 4.88 0.81 2.12 10.77

55 Activision 4.99 3.69 0.38 1.63 10.69

56 Take-Two Interactive 4.76 3.76 0.44 1.62 10.58

57 Nintendo 5.99 2.15 2.12 0.29 10.55

58 Nintendo 4.34 2.65 3.15 0.35 10.49

59 Nintendo 5.08 3.11 1.25 0.98 10.42

60 Ubisoft 6.05 3.15 0.00 1.07 10.27

61 Activision 6.72 2.63 0.04 0.82 10.21

62 Microsoft Game Studios 7.03 1.98 0.08 0.78 9.87

63 Nintendo 5.55 1.94 2.23 0.15 9.87

64 Nintendo 3.66 3.07 2.47 0.63 9.83

65 Microsoft Game Studios 6.63 2.36 0.04 0.73 9.76

66 Sony Computer Entertainment 3.01 2.47 3.28 0.96 9.72

67 Activision 4.09 3.73 0.38 1.38 9.58

68 Ubisoft 5.84 2.89 0.01 0.78 9.52

69 Sony Computer Entertainment 3.88 3.42 1.69 0.50 9.49

70 Activision 5.91 2.38 0.13 0.90 9.32

71 Nintendo 4.36 1.71 3.00 0.23 9.30

72 Microsoft Game Studios 5.58 2.83 0.02 0.77 9.20

73 Nintendo 2.01 2.32 4.36 0.41 9.10

74 Nintendo 4.46 1.88 1.98 0.70 9.02

75 Bethesda Softworks 5.03 2.86 0.10 0.85 8.84

76 Nintendo 3.54 1.24 3.81 0.18 8.77

77 Electronic Arts 1.11 6.06 0.06 1.26 8.49

78 Nintendo 1.79 3.53 2.49 0.68 8.49

79 Microsoft Game Studios 6.82 1.53 0.05 0.08 8.48

80 Nintendo 3.81 2.30 1.58 0.73 8.42

81 Nintendo 2.91 1.86 3.14 0.43 8.34

82 Electronic Arts 1.06 5.05 0.13 2.01 8.25

83 Electronic Arts 0.98 6.42 0.00 0.71 8.11

84 Nintendo 5.80 2.01 0.13 0.15 8.09

85 Sega 2.58 3.90 0.66 0.91 8.05

86 Sony Computer Entertainment 2.91 2.07 2.73 0.33 8.04

87 SquareSoft 2.28 1.72 3.63 0.23 7.86

88 Nintendo 2.82 1.78 2.69 0.55 7.84

89 Atari 7.28 0.45 0.00 0.08 7.81

90 Take-Two Interactive 2.90 2.83 0.24 1.75 7.72

[ reached 'max' / getOption("max.print") -- omitted 16507 rows ]

> RNremove(GenreSalesGlobal)

genre globalSales

1 Action 1750.16

2 Adventure 238.81

3 Fighting 448.94

4 Misc 809.30

5 Platform 831.04

6 Puzzle 244.42

7 Racing 731.77

8 Role-Playing 927.26

9 Shooter 1036.84

10 Simulation 391.91

11 Sports 1247.80

12 Strategy 174.86

> #converting genre column from factor to character

> i <- sapply(GenreSalesGlobal, is.factor)

> GenreSalesGlobal[i] <- lapply(GenreSalesGlobal[i], as.character)

> str(GenreSalesGlobal)

'data.frame': 12 obs. of 2 variables:

$ genre : chr "Action" "Adventure" "Fighting" "Misc" ...

$ globalSales: num 1750 239 449 809 831 ...

> #analysis of genre sales by market

> statsVG<-function(x){

+ a<-mean(x)

+ b<-median(x)

+ c<-min(x)

+ d<-max(x)

+ e<-sd(x)

+ f<-quantile(x,probs = 0.05)

+ g<-quantile(x,probs = 0.95)

+ cat("mean:" ,a, "median:" ,b, "min:" ,c, "max" ,d, "sd:",e,"quantile(0.05-0.95):",f,"--",g)

+

+ }

> statsVG(genreSalesTotal$naSales)

mean: 362.6217 median: 343.35 min: 68.7 max 877.83 sd: 245.5669 quantile(0.05-0.95): 89.105 -- 748.0465> statsVG(genreSalesTotal$euSales)

mean: 200.4258 median: 194.845 min: 45.34 max 525 sd: 142.7231 quantile(0.05-0.95): 48.332 -- 427.5565> statsVG(genreSalesTotal$jpSales)

mean: 107.2708 median: 75.525 min: 38.28 max 352.31 sd: 86.58279 quantile(0.05-0.95): 44.429 -- 246.512> statsVG(genreSalesTotal$otherSales)

mean: 65.77417 median: 55.6 min: 11.36 max 187.38 sd: 52.66799 quantile(0.05-0.95): 12.0145 -- 153.9015> #creating a bar graph to compare genre success based on region

> gGlobal<-ggplot(GenreSalesGlobal, aes(x = reorder(genre, -globalSales),y=globalSales, fill = genre)) + geom\_bar(stat = "identity")

> gGlobal<-gGlobal + labs(y= 'Units Sales(millions)', x ='Video Game Genres') + ggtitle('Video Game Units Sold by Genre')

> gGlobal<-gGlobal + theme(axis.text.x = element\_text(angle = 90, hjust = 0.5))

> gGlobal

> #create a bar graph to show sales by genre across markets

> #make row names into a column called genre

> genreSalesTotal$genre<-rownames(genreSalesTotal)

> rownames(genreSalesTotal)<-NULL

> # compare genre performance across markets

> market<- c(rep("NaSales" , 12) , rep("EUSales" , 12) , rep("JPSales" , 12) , rep("OtherSales" , 12) )

> genre1 <- rep(c("Action" , "Adventure","Fighting", "Misc", "Platform","Puzzle","Racing","Role-Playing",

+ "Shooter", "Simulation","sports", "Strategy") , 4)

> Unit <- c(877.83,105.80,223.59,410.24,447.05,123.78,359.42,327.28,582.60,183.31,683.35,68.70,525.00,

+ 64.13,101.32,215.98,201.63,50.78,238.39,188.06,313.27,113.38,376.85,45.34,159.95,52.07,

+ 87.35,107.76,130.77,57.31,56.69,352.31,38.28,63.70,135.37,49.46,187.38,16.81,36.68,75.32,

+ 51.59,12.55,77.27,59.61,102.69,31.52,134.97,11.36)

> data <- data.frame(market,genre1,Unit)

> marketGenre<-ggplot(data, aes(fill=genre1,y=Unit, x=market,col= 'black')) +

+ geom\_bar(position="dodge", stat="identity") + scale\_color\_brewer(type = 'qual')

> marketGenre

> #TOP 3 GENRES:

> #EU: 1.Action, 2.Sports, and 3.Shooters

> #JP: 1.Role-Playing, 2.Action, and 3.Sports

> #NA: 1.Action, 2.Sports, and 3.Shooters

> #Other:

> #compare sales over time

> #Create matrix of total sales in each market by year. NA'S included

> NAsalesByDate<-tapply(VG$NA\_Sales,list(VG$Year,VG$Genre),sum)

> JPsalesByDate<-tapply(VG$JP\_Sales,list(VG$Year,VG$Genre),sum)

> EUsalesByDate<-tapply(VG$EU\_Sales,list(VG$Year,VG$Genre),sum)

> OthersalesByDate<-tapply(VG$Other\_Sales,list(VG$Year,VG$Genre),sum)

> #change NA'S to zeros

> NAsalesByDate[is.na(NAsalesByDate)]<-0

> JPsalesByDate[is.na(JPsalesByDate)]<-0

> EUsalesByDate[is.na(EUsalesByDate)]<-0

> OthersalesByDate[is.na(NAsalesByDate)]<-0

> #parce global data

> VG<-na.omit(VG)

> gGenreYear<-sqldf("SELECT sum(Global\_Sales), ")

Error: incomplete input

> g80<-sqldf("SELECT sum(Global\_Sales),Year FROM VG WHERE Year<1990 group by Year")

> g80

sum(Global\_Sales) Year

1 11.38 1980

2 35.68 1981

3 28.88 1982

4 16.80 1983

5 50.35 1984

6 53.95 1985

7 37.08 1986

8 21.70 1987

9 47.21 1988

10 73.45 1989

> g90<-sqldf("SELECT sum(Global\_Sales),Year FROM VG WHERE Year<2000 AND Year>1989 group by Year")

> g90

sum(Global\_Sales) Year

1 49.37 1990

2 32.23 1991

3 76.14 1992

4 45.99 1993

5 79.22 1994

6 88.11 1995

7 199.15 1996

8 201.07 1997

9 256.33 1998

10 251.12 1999

> g00<-sqldf("SELECT sum(Global\_Sales),Year FROM VG WHERE Year<2010 AND Year>1999 group by Year")

> g00

sum(Global\_Sales) Year

1 201.63 2000

2 331.49 2001

3 394.97 2002

4 357.61 2003

5 418.85 2004

6 459.41 2005

7 437.78 2006

8 610.44 2007

9 678.49 2008

10 667.10 2009

> g10<-sqldf("SELECT sum(Global\_Sales),Year FROM VG WHERE Year<2021 AND Year>2009 group by Year")

> g10

sum(Global\_Sales) Year

1 600.36 2010

2 515.93 2011

3 363.30 2012

4 367.98 2013

5 337.10 2014

6 264.26 2015

7 70.87 2016

8 0.05 2017

9 0.29 2020

> gTotal<-sqldf("SELECT sum(Global\_Sales), Year FROM VG group by Year")

> gTotal

sum(Global\_Sales) Year

1 11.38 1980

2 35.68 1981

3 28.88 1982

4 16.80 1983

5 50.35 1984

6 53.95 1985

7 37.08 1986

8 21.70 1987

9 47.21 1988

10 73.45 1989

11 49.37 1990

12 32.23 1991

13 76.14 1992

14 45.99 1993

15 79.22 1994

16 88.11 1995

17 199.15 1996

18 201.07 1997

19 256.33 1998

20 251.12 1999

21 201.63 2000

22 331.49 2001

23 394.97 2002

24 357.61 2003

25 418.85 2004

26 459.41 2005

27 437.78 2006

28 610.44 2007

29 678.49 2008

30 667.10 2009

31 600.36 2010

32 515.93 2011

33 363.30 2012

34 367.98 2013

35 337.10 2014

36 264.26 2015

37 70.87 2016

38 0.05 2017

39 0.29 2020

40 99.99 N/A

> nintendoGlobal<-sqldf("SELECT sum(Global\_Sales),Publisher,Year FROM VG WHERE Publisher = 'Nintendo' group by Year")

> nintendoGlobal

sum(Global\_Sales) Publisher Year

1 10.96 Nintendo 1983

2 45.55 Nintendo 1984

3 49.95 Nintendo 1985

4 16.17 Nintendo 1986

5 11.95 Nintendo 1987

6 36.44 Nintendo 1988

7 63.87 Nintendo 1989

8 35.47 Nintendo 1990

9 15.97 Nintendo 1991

10 38.13 Nintendo 1992

11 20.04 Nintendo 1993

12 25.00 Nintendo 1994

13 16.71 Nintendo 1995

14 73.69 Nintendo 1996

15 25.82 Nintendo 1997

16 48.40 Nintendo 1998

17 65.31 Nintendo 1999

18 34.02 Nintendo 2000

19 45.41 Nintendo 2001

20 48.32 Nintendo 2002

21 38.17 Nintendo 2003

22 60.66 Nintendo 2004

23 127.37 Nintendo 2005

24 122.86 Nintendo 2006

25 104.16 Nintendo 2007

26 91.14 Nintendo 2008

27 128.86 Nintendo 2009

28 61.07 Nintendo 2010

29 51.53 Nintendo 2011

30 56.50 Nintendo 2012

31 52.78 Nintendo 2013

32 48.67 Nintendo 2014

33 27.08 Nintendo 2015

34 3.46 Nintendo 2016

35 2.13 Nintendo N/A

> nintendoNA<-sqldf("SELECT sum(NA\_Sales),Publisher,Year FROM VG WHERE Publisher = 'Nintendo' group by Year")

> nintendoNA

sum(NA\_Sales) Publisher Year

1 2.32 Nintendo 1983

2 32.57 Nintendo 1984

3 32.48 Nintendo 1985

4 6.73 Nintendo 1986

5 5.13 Nintendo 1987

6 21.23 Nintendo 1988

7 39.81 Nintendo 1989

8 19.75 Nintendo 1990

9 6.95 Nintendo 1991

10 18.48 Nintendo 1992

11 9.53 Nintendo 1993

12 10.96 Nintendo 1994

13 6.70 Nintendo 1995

14 34.54 Nintendo 1996

15 15.80 Nintendo 1997

16 23.42 Nintendo 1998

17 34.18 Nintendo 1999

18 15.00 Nintendo 2000

19 23.84 Nintendo 2001

20 25.05 Nintendo 2002

21 20.15 Nintendo 2003

22 28.30 Nintendo 2004

23 45.05 Nintendo 2005

24 48.86 Nintendo 2006

25 42.77 Nintendo 2007

26 38.83 Nintendo 2008

27 53.15 Nintendo 2009

28 24.18 Nintendo 2010

29 20.24 Nintendo 2011

30 20.84 Nintendo 2012

31 19.00 Nintendo 2013

32 18.01 Nintendo 2014

33 9.25 Nintendo 2015

34 1.16 Nintendo 2016

35 1.12 Nintendo N/A

> nintendoJP<-sqldf("SELECT sum(JP\_Sales),Publisher,Year FROM VG WHERE Publisher = 'Nintendo' group by Year")

> nintendoJP

sum(JP\_Sales) Publisher Year

1 8.10 Nintendo 1983

2 10.36 Nintendo 1984

3 12.02 Nintendo 1985

4 7.45 Nintendo 1986

5 5.46 Nintendo 1987

6 8.13 Nintendo 1988

7 15.06 Nintendo 1989

8 8.08 Nintendo 1990

9 6.22 Nintendo 1991

10 12.66 Nintendo 1992

11 6.99 Nintendo 1993

12 9.00 Nintendo 1994

13 7.45 Nintendo 1995

14 21.14 Nintendo 1996

15 4.23 Nintendo 1997

16 11.32 Nintendo 1998

17 15.88 Nintendo 1999

18 10.53 Nintendo 2000

19 11.15 Nintendo 2001

20 11.90 Nintendo 2002

21 9.85 Nintendo 2003

22 17.47 Nintendo 2004

23 29.50 Nintendo 2005

24 32.97 Nintendo 2006

25 19.98 Nintendo 2007

26 19.96 Nintendo 2008

27 29.82 Nintendo 2009

28 16.30 Nintendo 2010

29 13.43 Nintendo 2011

30 18.56 Nintendo 2012

31 14.39 Nintendo 2013

32 16.12 Nintendo 2014

33 8.57 Nintendo 2015

34 1.17 Nintendo 2016

35 0.43 Nintendo N/A

> nintendoEU<-sqldf("SELECT sum(EU\_Sales),Publisher,Year FROM VG WHERE Publisher = 'Nintendo' group by Year")

> nintendoEU

sum(EU\_Sales) Publisher Year

1 0.46 Nintendo 1983

2 1.95 Nintendo 1984

3 4.54 Nintendo 1985

4 1.73 Nintendo 1986

5 1.18 Nintendo 1987

6 6.15 Nintendo 1988

7 7.60 Nintendo 1989

8 6.42 Nintendo 1990

9 2.47 Nintendo 1991

10 6.08 Nintendo 1992

11 3.09 Nintendo 1993

12 4.19 Nintendo 1994

13 2.21 Nintendo 1995

14 16.31 Nintendo 1996

15 5.22 Nintendo 1997

16 12.27 Nintendo 1998

17 13.74 Nintendo 1999

18 6.42 Nintendo 2000

19 8.84 Nintendo 2001

20 9.89 Nintendo 2002

21 7.08 Nintendo 2003

22 12.43 Nintendo 2004

23 42.69 Nintendo 2005

24 31.33 Nintendo 2006

25 32.81 Nintendo 2007

26 25.13 Nintendo 2008

27 36.18 Nintendo 2009

28 16.52 Nintendo 2010

29 14.38 Nintendo 2011

30 13.97 Nintendo 2012

31 16.21 Nintendo 2013

32 11.37 Nintendo 2014

33 7.49 Nintendo 2015

34 0.93 Nintendo 2016

35 0.44 Nintendo N/A

> nintendoOther<-sqldf("SELECT sum(Other\_Sales),Publisher,Year FROM VG WHERE Publisher = 'Nintendo' group by Year")

> nintendoOther

sum(Other\_Sales) Publisher Year

1 0.08 Nintendo 1983

2 0.67 Nintendo 1984

3 0.91 Nintendo 1985

4 0.26 Nintendo 1986

5 0.18 Nintendo 1987

6 0.93 Nintendo 1988

7 1.40 Nintendo 1989

8 1.22 Nintendo 1990

9 0.33 Nintendo 1991

10 0.91 Nintendo 1992

11 0.43 Nintendo 1993

12 0.85 Nintendo 1994

13 0.35 Nintendo 1995

14 1.70 Nintendo 1996

15 0.57 Nintendo 1997

16 1.39 Nintendo 1998

17 1.51 Nintendo 1999

18 2.07 Nintendo 2000

19 1.58 Nintendo 2001

20 1.48 Nintendo 2002

21 1.09 Nintendo 2003

22 2.46 Nintendo 2004

23 10.13 Nintendo 2005

24 9.70 Nintendo 2006

25 8.60 Nintendo 2007

26 7.22 Nintendo 2008

27 9.71 Nintendo 2009

28 4.07 Nintendo 2010

29 3.48 Nintendo 2011

30 3.13 Nintendo 2012

31 3.18 Nintendo 2013

32 3.17 Nintendo 2014

33 1.77 Nintendo 2015

34 0.20 Nintendo 2016

35 0.14 Nintendo N/A

> nintendoTotal<-as.data.frame(nintendoNA,nintendoEU,nintendoJP,nintendoOther)

Error in as.data.frame.data.frame(nintendoNA, nintendoEU, nintendoJP, :

invalid 'row.names', length 3 for a data frame with 35 rows

> #rename axis lables

> colnames(gTotal)<-c("Global\_Sales","Year")

> colnames(g80)<-c("Global\_Sales","Year")

> colnames(g90)<-c("Global\_Sales","Year")

> colnames(g00)<-c("Global\_Sales","Year")

> colnames(g10)<-c("Global\_Sales","Year")

> #plot the time lines of global sales

> g80P<-ggplot(g80,aes(y=Global\_Sales, x=Year, group = 1))+geom\_line()+geom\_point(data = g80)

> g80P<-g80P + ggtitle("Global Sales in the 1980's")

> g80P

> g90P<-ggplot(g90,aes(y=Global\_Sales, x=Year, group = 1))+geom\_line(color="orchid")+geom\_point(color="blue",data = g90)

> g90P<-g90P + ggtitle("Global Sales in the 1990's") + theme(plot.title=element\_text(hjust=.5))+ylab("Global Sales in Millions")

> g90P

> g00P<-ggplot(g00,aes(y=Global\_Sales, x=Year, group = 1))+geom\_line(color="orchid")+geom\_point(color="blue",data = g00)

> g00P<-g00P + ggtitle("Global Sales in the 2000's") + theme(plot.title=element\_text(hjust=.5))+ylab("Global Sales in Millions")

> g00P

> g10P<-ggplot(g10,aes(y=Global\_Sales, x=Year, group = 1))+geom\_line(color="orchid")+geom\_point(color="blue",data = g00)

> g10P<-g10P + ggtitle("Global Sales in the 2010's") + theme(plot.title=element\_text(hjust=.5))+ylab("Global Sales in Millions")

> g10P

> gTotalP<-ggplot(gTotal,aes(y=Global\_Sales, x=Year, group = 1))+geom\_line(color="orchid")+geom\_point(color="blue",data = gTotal)

> gTotalP<-gTotalP + ggtitle("Global Sales from 1980 to 2020") + theme(plot.title=element\_text(hjust=.5))+ylab("Global Sales in Millions")

> gTotalP

>

> #mike's code#

> devtools::install\_github("robj")

Error in parse\_repo\_spec(repo) : Invalid git repo specification: 'robj'

> #### Call esquisse ggplot plug-in ######

> esquisse::esquisser()

Error in loadNamespace(name) : there is no package called ‘esquisse’

> ###Global Sales by Genre – Fill Platform & Color Platform

> ggplot(VG) +

+ aes(x = Genre, fill = Platform, colour = Platform, weight = Global\_Sales) +

+ geom\_bar() +

+ scale\_fill\_hue() +

+ scale\_color\_hue() +

+ theme\_minimal()

> ###Global Sales by Publisher – Fill Platform & faceted by year \* (slice by decade?)

> VG %>%

+ filter(Global\_Sales >= 9 & Global\_Sales <= 40.5) %>%

+ ggplot() +

+ aes(x = Publisher, fill = Platform, weight = Global\_Sales) +

+ geom\_bar() +

+ scale\_fill\_hue() +

+ labs(y = "Global Sales") +

+ coord\_flip() +

+ theme\_minimal() +

+ facet\_wrap(vars(Year))

> #Global Sales Platform V Publisher

> VG %>%

+ filter(Global\_Sales >= 8 & Global\_Sales <= 40.5) %>%

+ ggplot() +

+ aes(x = Platform, fill = Publisher, weight = Global\_Sales) +

+ geom\_bar() +

+ scale\_fill\_hue() +

+ labs(y = "Global Sales", title = "Global Sales ", subtitle = "Plaform V Publisher") +

+ theme\_minimal()

> #Genre Popularity – Sales over Time

>

> VG %>%

+ filter(Global\_Sales >= 8 & Global\_Sales <= 40.5) %>%

+ ggplot() +

+ aes(x = Year, fill = Genre, weight = Global\_Sales) +

+ geom\_bar() +

+ scale\_fill\_hue() +

+ labs(x = "Year", y = "Global Sales", title = "Genre Popularity", subtitle = "1984 - 2015") +

+ theme\_minimal()

> #Platform Popularity -

> VG %>%

+ filter(Global\_Sales >= 8 & Global\_Sales <= 40.5) %>%

+ ggplot() +

+ aes(x = Year, fill = Platform, weight = Global\_Sales) +

+ geom\_bar() +

+ scale\_fill\_hue() +

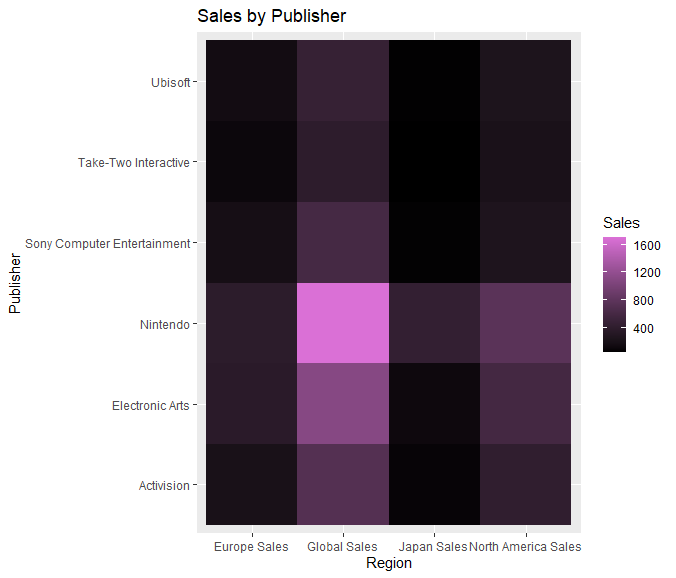
+ labs(x = "Year", y = "Global Sales", title = "Platform Popularity", subtitle = "1984 - 2015") +

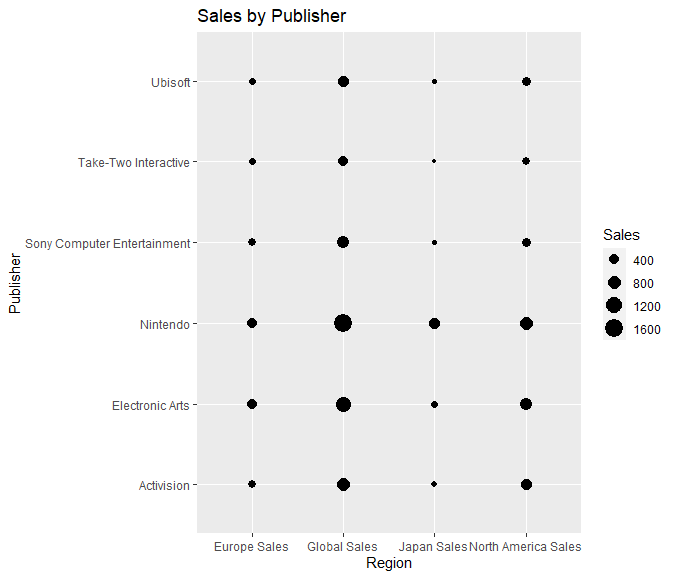
+ theme\_minimal()

**Appendix C**

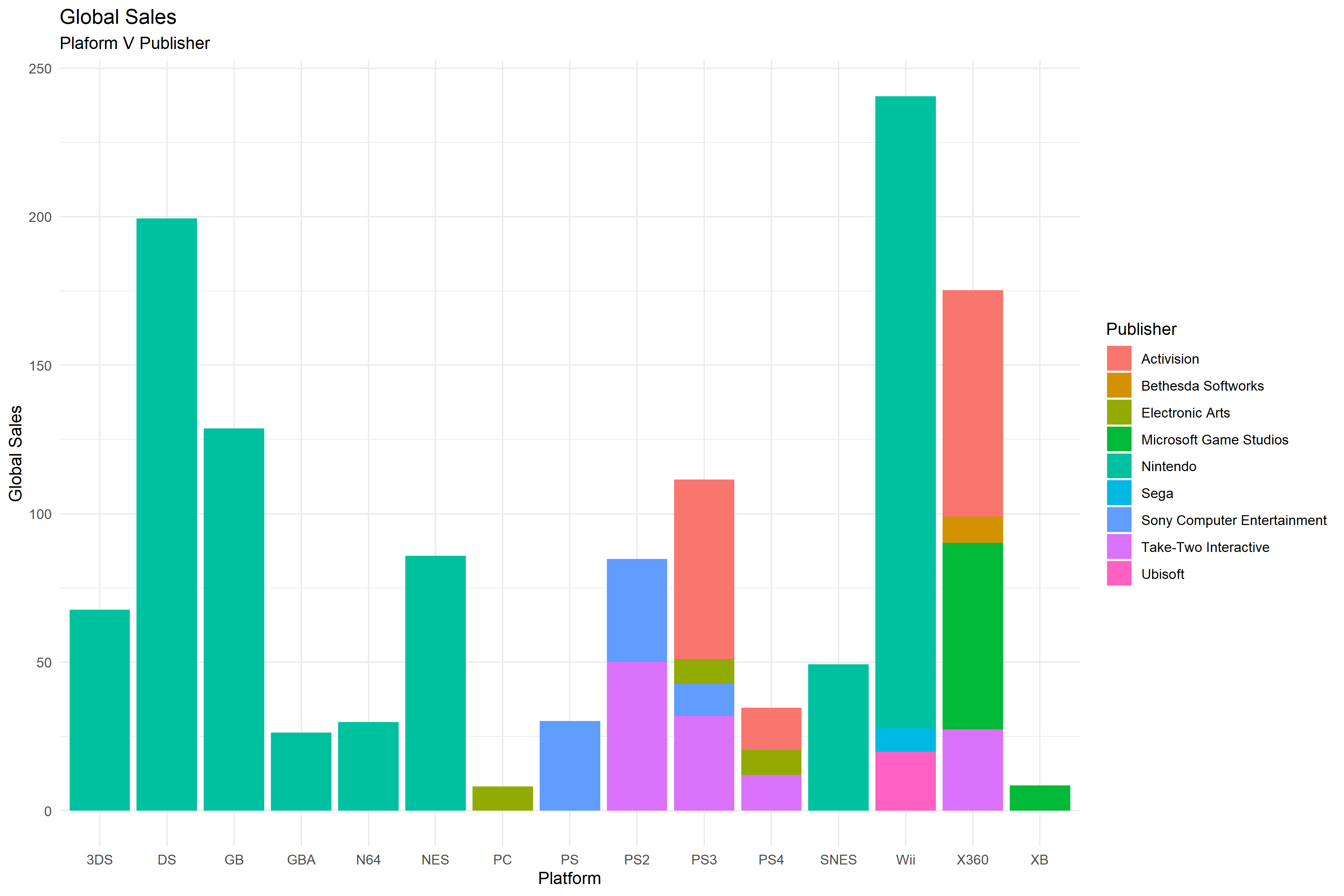
R Code – Visualizations

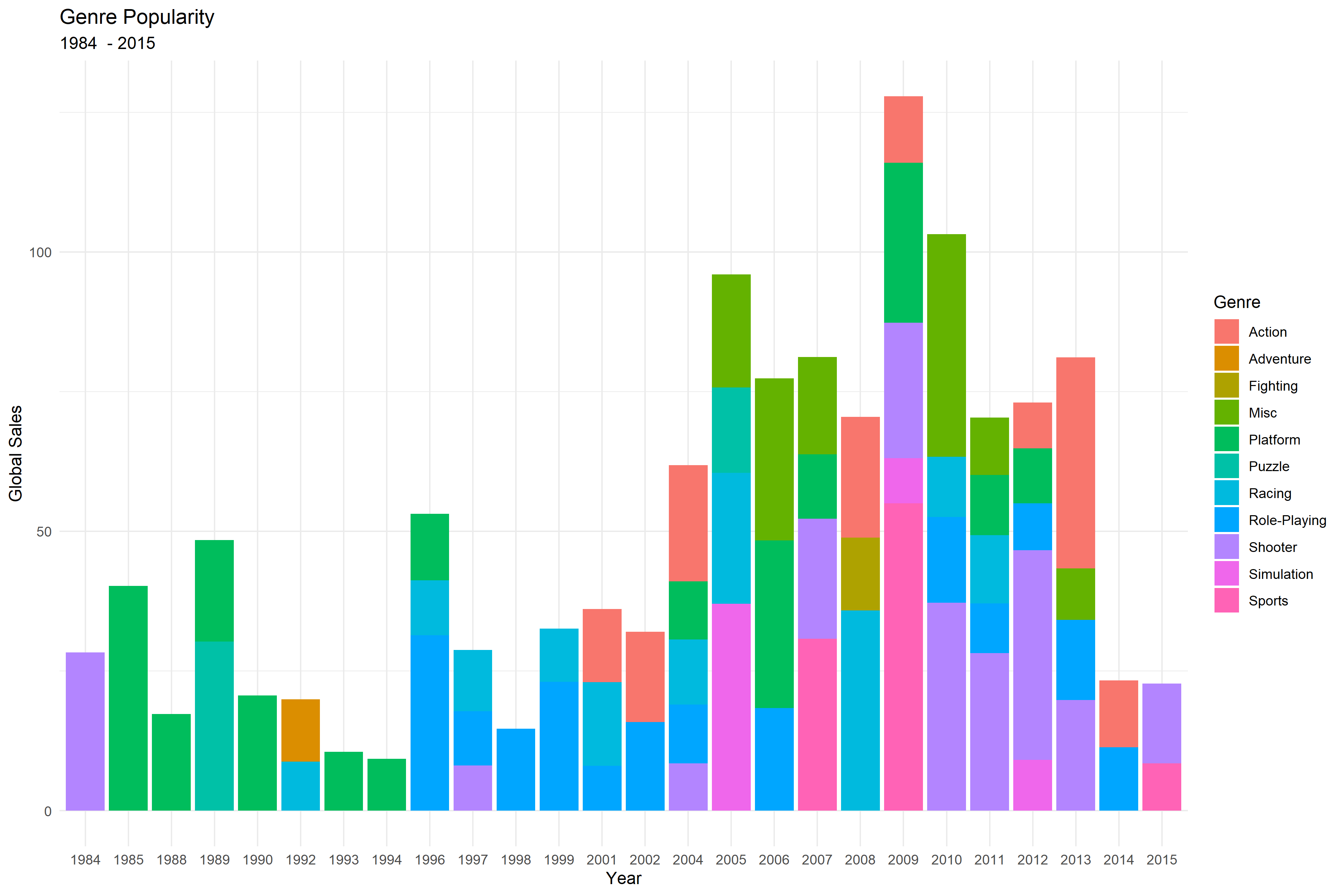
#Jennifer Plots

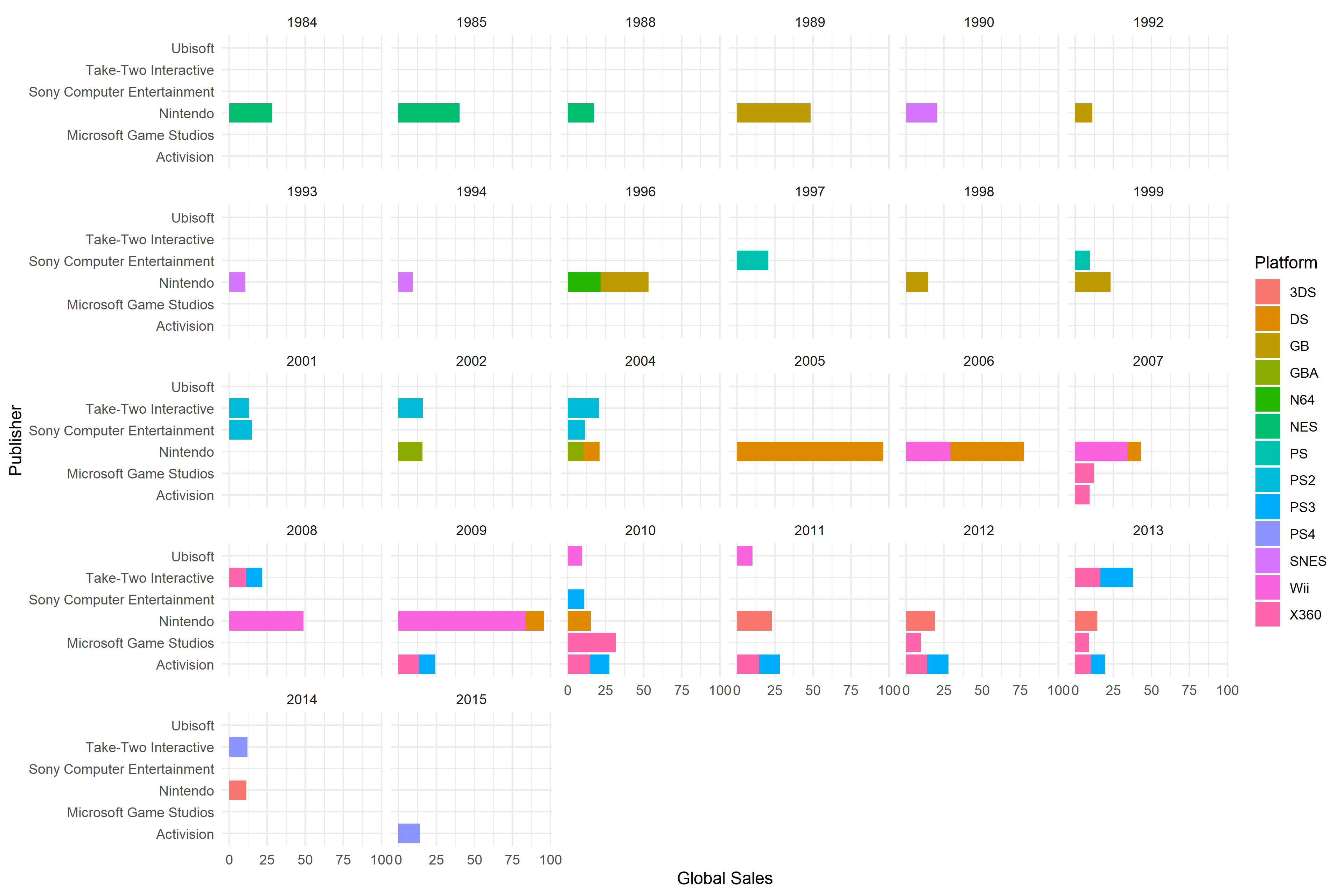


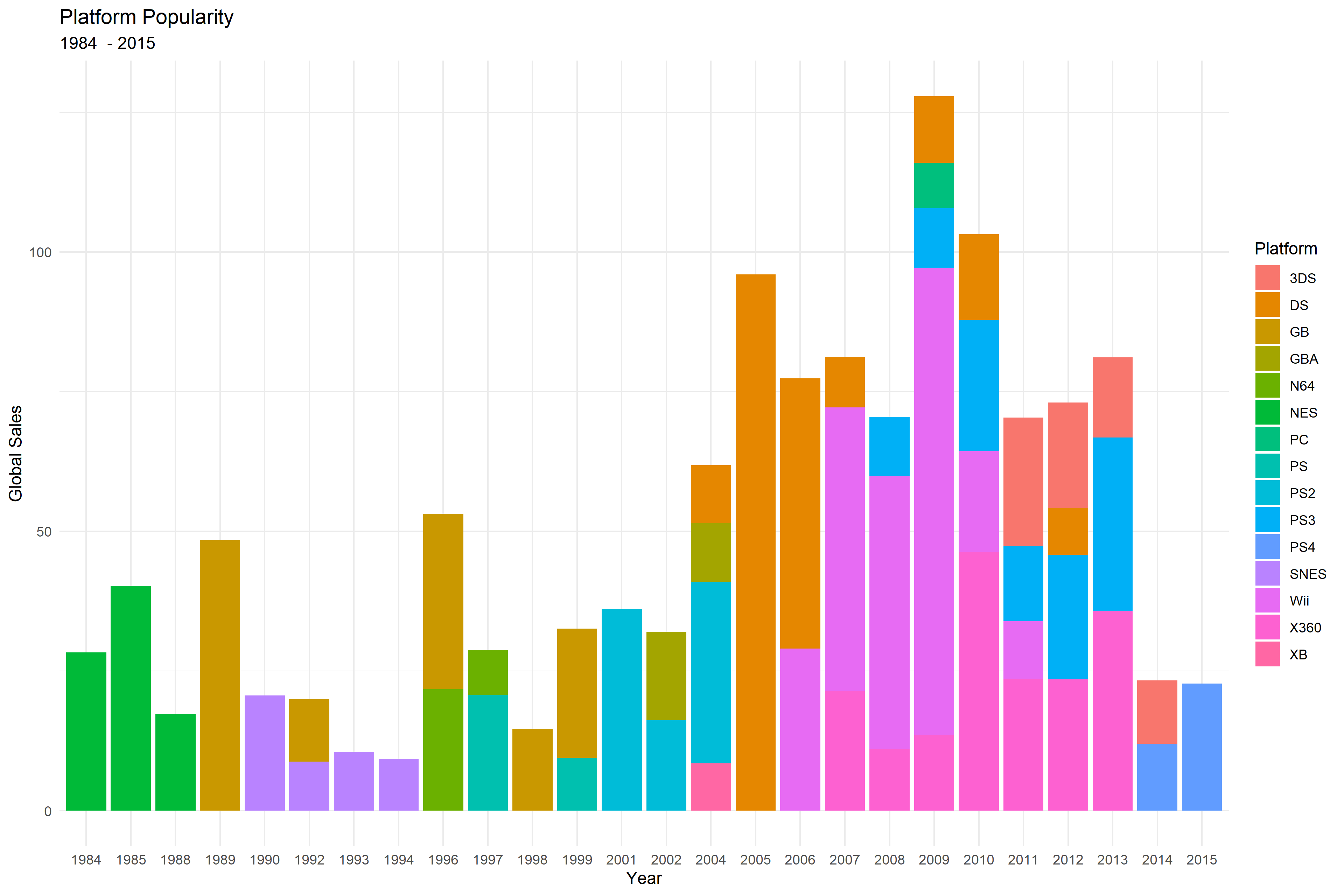


#Mike Visualizations

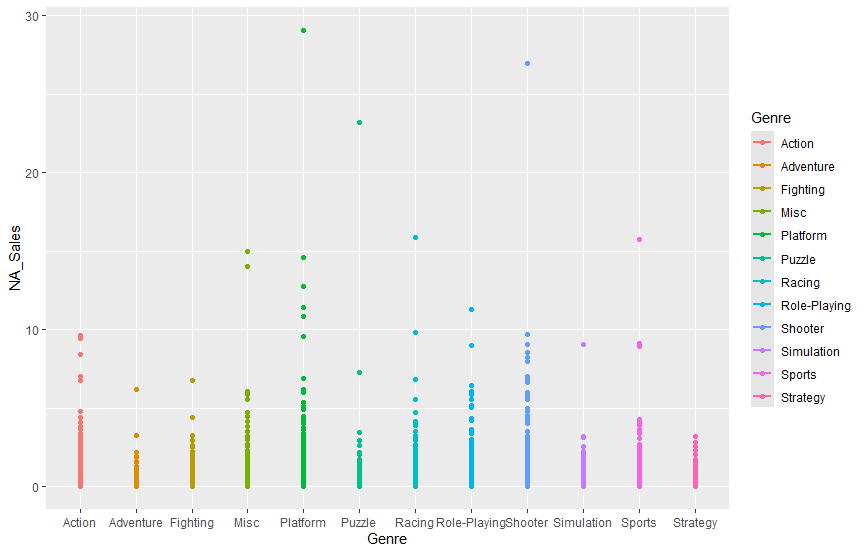




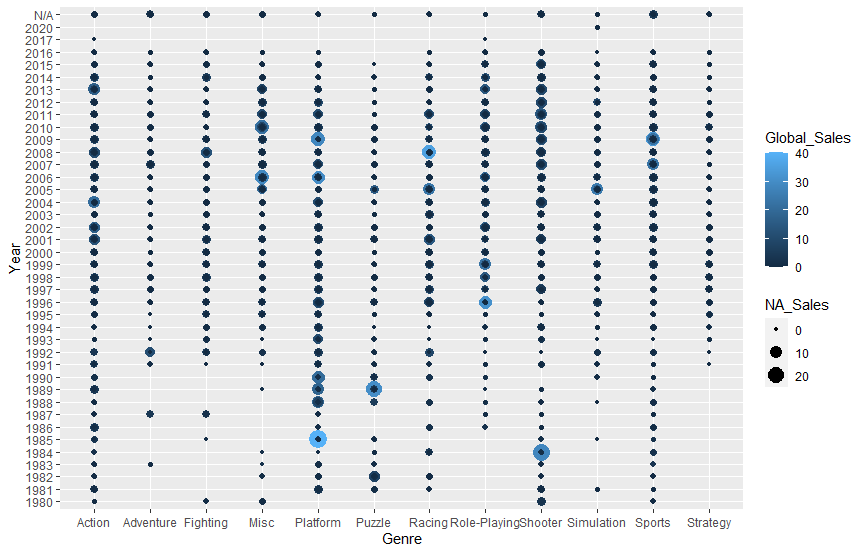


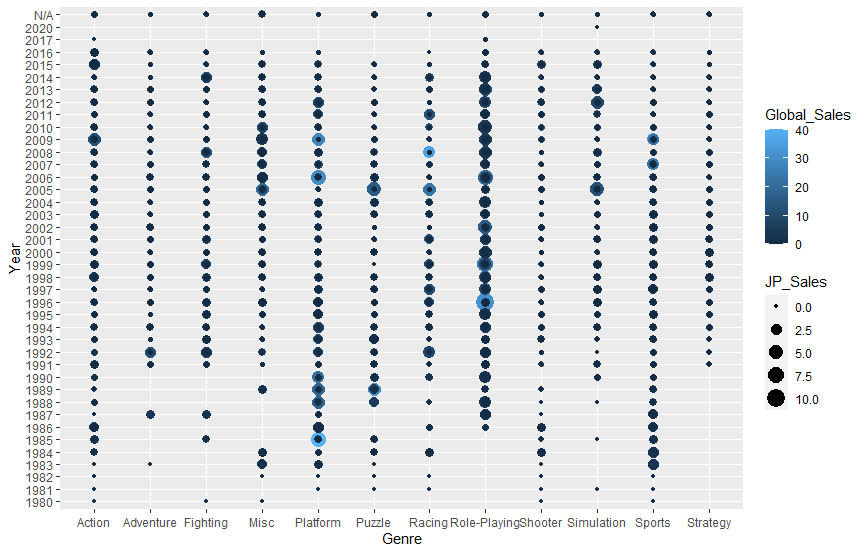


#Nick Visualization

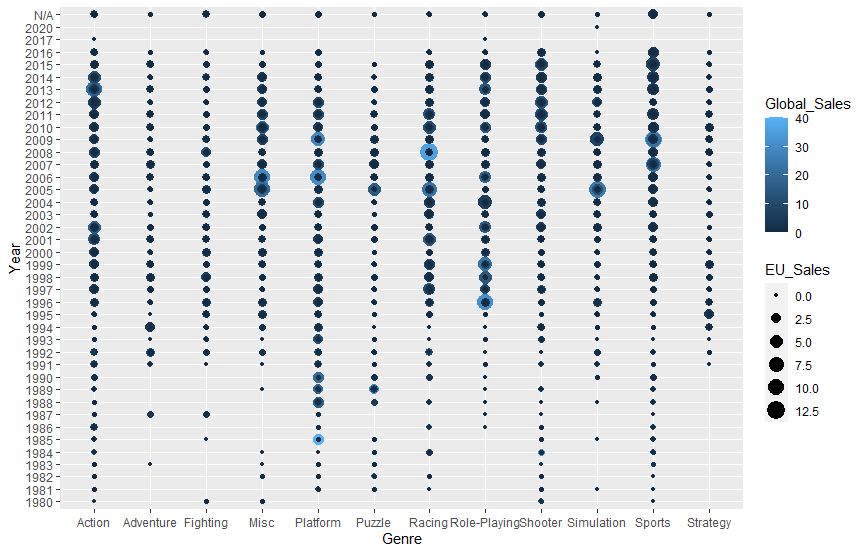


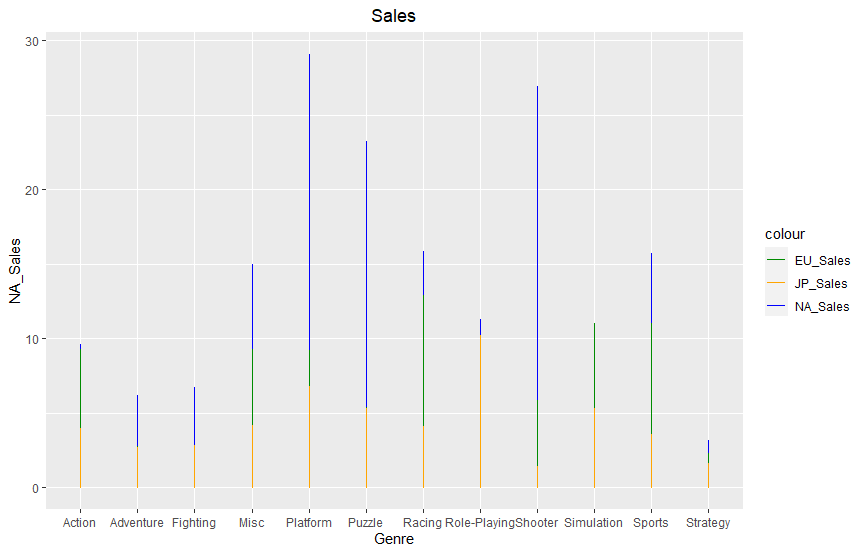
Alternative Version of the plot below.

 Typically, games that do well in NA are the ones that do well worldwide but there are exceptions in 1985 the platform genre sold a lot more games in other countries. The platform genre is the most common genre to have more sales outside of NA

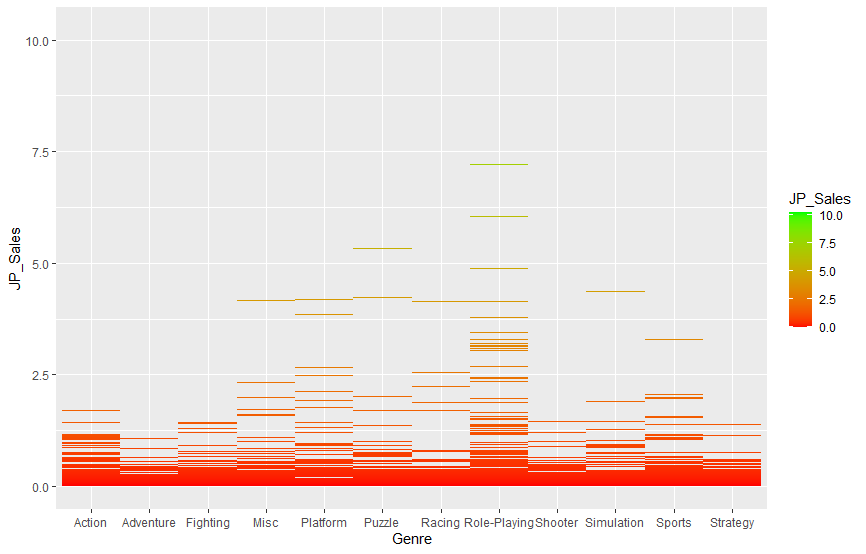


Japan almost always has a large portion of the role playing game sales even they don’t make up a large portion of global sales

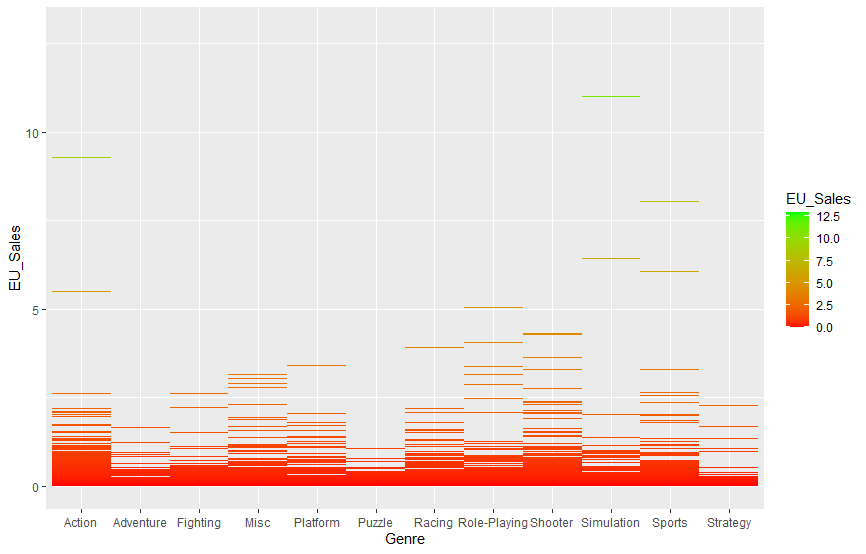




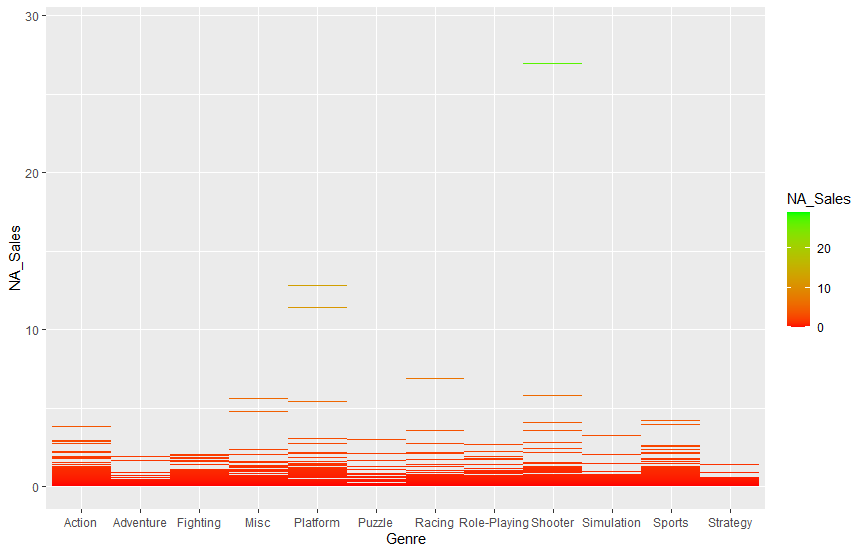
NA has the most sales in every genre except simulation which EU has the most in and NA only has a slight lead in the action genre.



Japanese sales are more concentrated in the roleplaying, platform, and puzzle games

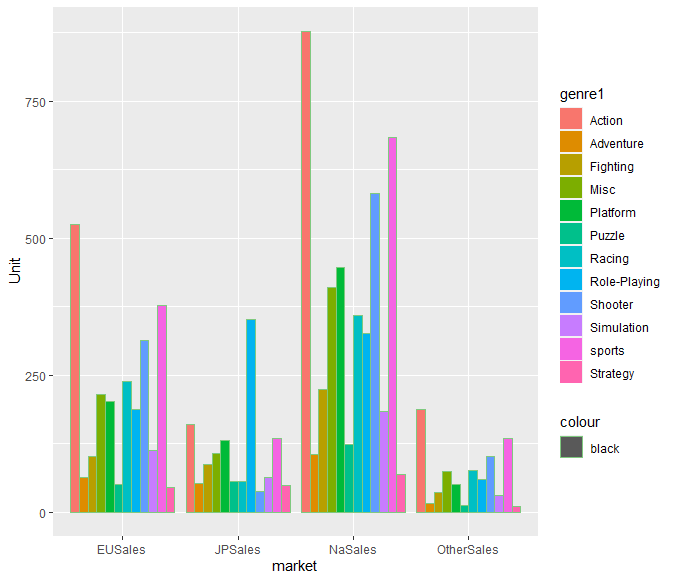


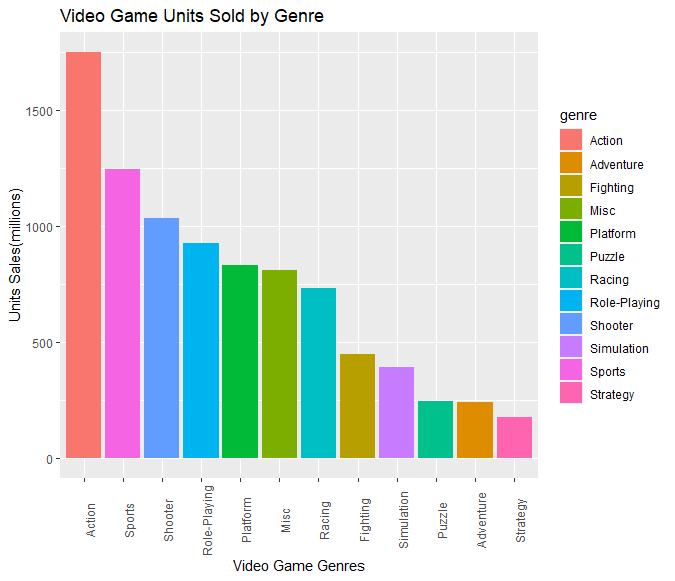
EU games are more centered around Sports, Role-playing and action games



NA Sales show the two highest performing genres are Platform and Shooter with racing as a third.

#Joshua Visualizations





#Joshua Line Charts

