**Video Game Sales**

**An Exploration of Types and Trends**

**Author: Jacob Walsh**

**PSU Email: jvw5761@psu.edu**

**Abstract**

Associations between the genre of a title and the platform it is released on are explored in this report. In addition, differences in global sales and critic ratings across genres and platforms are explored. Since there are over 20 platforms, titles were grouped into broader categories based on similarities. The first grouping classifies a title as either a console game or personal computer (PC) game. The second grouping then splits the console group into two categories – handheld and non handheld. The third grouping splits the console group by manufacturers - Nintendo, Playstation, Xbox, and the PC.

Chi-Squared tests and odds ratios were used to compare the genres and platform types. Categorical comparisons revealed significant associations between genre and platform when comparing the personal computer to the consoles. Significant associations were also found between handheld and non-handheld games, as well as the Playstation, Xbox, and Nintendo. The Playstation and Xbox showed the most similarities in titles released. Comparisons of these platforms in the same time frame (ie. PS3 vs. Xbox 360) are the only ones that did not have a significant association between genre and platform.

Linear models, ANOVA, and intervals using Tukey’s method were used to assess differences in sales and ratings by genre and platform. There was a significant positive association between critic ratings and sales when considering each platform. This didn’t necessarily mean that the highest rated games sold the most though. Games on the PC were consistently rated higher but sold less than games on the other consoles.

**Introduction**

Over the past few decades consoles for playing video games have become commonplace in houses across the country. Software and hardware advancements of the personal computer (PC) have led to many different machines being developed. Every year, the video game industry churns out hundreds of titles and sells hundreds of millions of units around the globe. Many consoles have come and gone, but a few modern machines have carved their niche in society and continue to evolve and thrive. Xbox (developed by Microsoft), Playstation (developed by Sony), Nintendo, and the personal computer are examples of the platforms that have the biggest hit games from a variety of publishers.

In this report, the following questions of interest are addressed:

* Is there any evidence of differences between sales in different regions of the world?
* Are there any associations when comparing the genre of a game and the platform on which it is released?
* Are there any associations in sales or titles when we consider more current popular platforms?
* Do critic ratings correlate with sales? Do critics tend to favor some publishers/genres/platforms over others?

A customer’s purchase choice may be impacted by the types and quality of games in addition to the cost of the machine. A gaming computer system will cost more than consoles and this report will show that the types of games released on the PC are generally different than those on the console. The results in this paper may also provide some insight to consumers that are trying to decide which type of console is the best fit for their gaming desires.

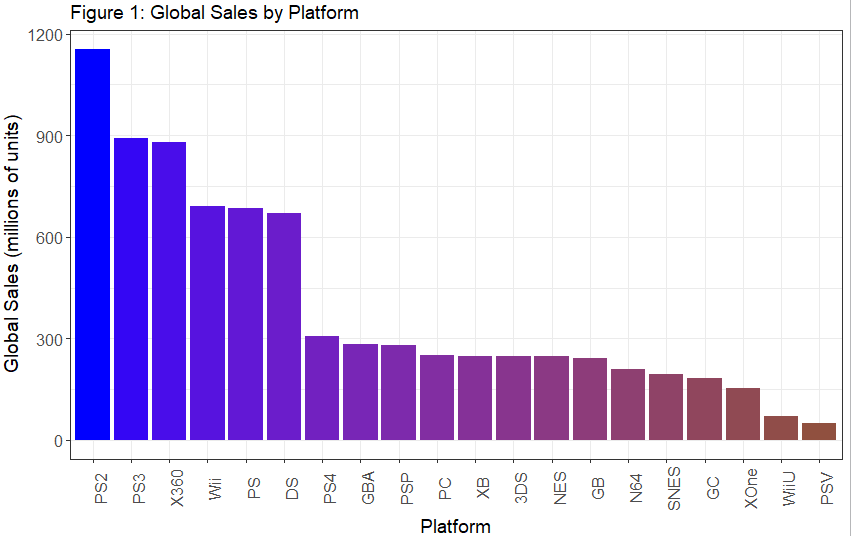
**Background**

VGChartz.com has data publicly available where total worldwide sales (in millions of units) are recorded with categorical information of genre, publisher, platform, and year of release for each title released. For all platforms that are no longer manufactured, sales are recorded as total units shipped from the manufacturers. Critic ratings from the site metacritic.com were added to the data for over 6,000 of the titles. Associations among titles, sales, critic ratings, genres and platforms of historical sales and not newly released titles, which have a greater margin of error because new sales are estimated, will be explored in this report.

Records for over 8.7 billion units of sale from 1983 to 2016 are available in this sample and classified by genre, platform, and year of release. Based on the available data, Nintendo published 7 of the top 10 selling games.

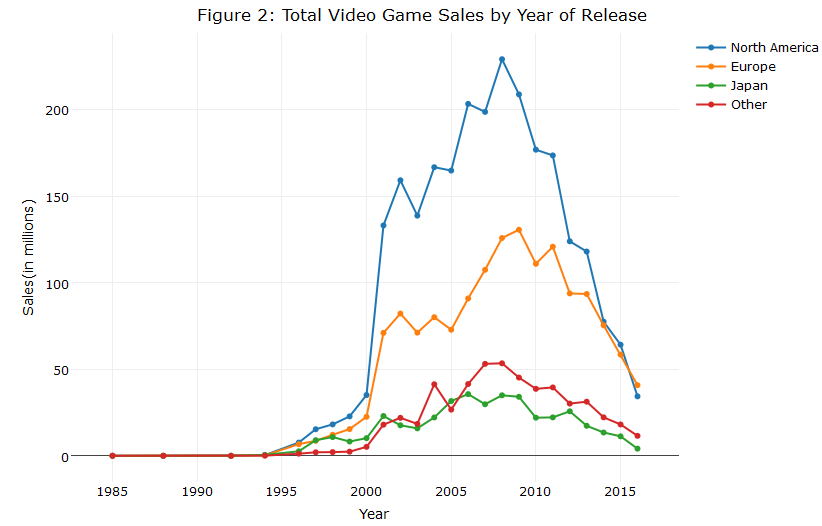
|  |  |  |
| --- | --- | --- |
| **Table 1: Best Selling Games** |  |  |
| **Title** | **Publisher** | **Global Sales (in millions of units)** |
| Wii Sports | Nintendo | 82.53 |
| Grand Theft Auto V | Take-Two Interactive | 56.57 |
| Super Mario Bros. | Nintendo | 45.31 |
| Tetris | Nintendo | 35.84 |
| Mario Kart Wii | Nintendo | 35.52 |
| Wii Sports Resort | Nintendo | 32.77 |
| Pokemon Red/Pokemon Blue | Nintendo | 31.37 |
| Call of Duty: Black Ops | Activision | 30.82 |
| Call of Duty: Modern Warfare 3 | Activision | 30.59 |
| New Super Mario Bros. | Nintendo | 29.8 |

Nintendo tops the charts for individual title sales (which can be partially attributed to bundled sales: ie #1 Wii Sports –Packaged with Wii until 2011 and #3 Super Mario Bros – Packaged with NES), but as can be seen in Figure 1, Playstation 2 has sold the most video games worldwide. The highest selling Nintendo platform is the Wii, coming in at fourth place.



**Research Question 1: Is there any evidence of differences between sales in different regions of the world?**

Sales records are available for North America, Europe, Japan, and other regions. As Figure 2 shows, from 1995 to 2009 the video game sales market went through a rapid growth period. In this time frame, Sony released the first Playstation (December of 1994) and Nintendo released the Nintendo 64 (June of 1996).



While North America and Europe’s sales have followed a similar trend Japan’s sales history lacks the rapid growth spurt after the year 2000 that is shown in the North America and Europe. Japan’s sales also have a lot less variation from year to year. As can be seen in Table 2, sales in North America and Europe had the strongest correlation of the regions. Japan was weakly correlated with the others. Clearly, sales in the regions differ with North America selling the most, then Europe, then Japan. In North America and in Europe, the best selling console was the Playstation 2. Japan differs in that the handheld gaming devices are more popular. Nintendo Gameboy and Nintendo DS games were the sold the most in Japan.

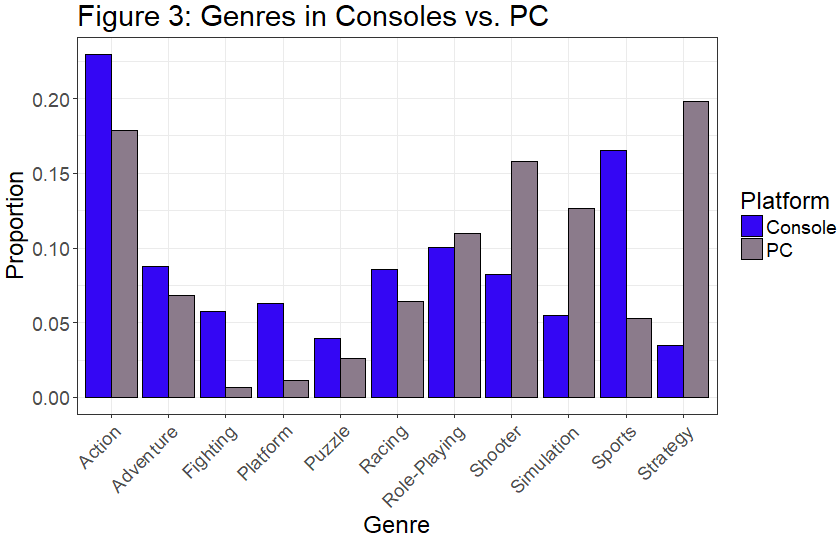
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Table 2: Correlation of Sales by Region** | |  |  |  |
|  | **N. America Sales** | **EuropeSales** | **Japan Sales** | **Other Sales** |
| N. America Sales | 1 | 0.768008 | 0.4699036 | 0.6342168 |
| Europe Sales | 0.768008 | 1 | 0.433727 | 0.7156443 |
| Japan Sales | 0.4699036 | 0.433727 | 1 | 0.2889267 |
| Other Sales | 0.6342168 | 0.7156443 | 0.2889267 | 1 |

**Research Question 2: Are there any associations when comparing the genre of a game and the platform on which it is released?**

Since over 20 platforms were included in the data set, grouping was done based on similarities of the platforms. Chi-squared testing and odds ratio analysis was done on three different groups. The first grouping classifies a title as either a console game or personal computer (PC) game. The second grouping then splits the console group titles into two categories – handheld and non handheld. The third grouping splits the console group by manufacturers - Nintendo, Playstation, Xbox, and the PC.

**Results for Genre Comparisons: Consoles and the PC**

Since the console and PC are different ways of gaming, one would suspect the types of games released on each may differ. A test for association between genre and type of platform yields a test statistic of ꭓ2 = 833.37 with 10 degrees of freedom. There is very strong evidence against the hypothesis that titles released on consoles and PC’s are independent of genre. Figure 3 below displays the proportion of titles released on the console and PC by genre.



Both types of platforms release a large proportion of action games. The console systems tended to release fighting, platform and sports games, while the PC favored shooter, simulation, and strategy games.

A summary of the odds-likelihood of a game in given genre to be released on a platform is shown in table 3. Each odds-ratio can be interpreted as the increased likelihood of observing the given genre on a PC when compared to action games on the console. If a title is known to be for the PC and not on a console, it is about 7 times more likely to be a strategy game than an action game and 2-3 times less likely to be a shooter or simulation game. Games were also about 7 times less likely to be fighting games, 4 times less likely to be platform games, and half as likely to be sport’s games if they were released on the PC.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Table 3: Significant Odds Ratios when Comparing a Given Genre on PC vs. Console** | | | | | |  |
| **Genre** | **Console Titles** | **PC Titles** | **Odds** | **95% Lower Bound** | **95% Upper Bound** | |
| Fighting | 782 | 6 | 0.1412667 | 0.06234841 | 0.320077 | |
| Platform | 852 | 11 | 0.2377106 | 0.12858597 | 0.4394439 | |
| Shooter | 1122 | 150 | 2.4614659 | 1.95589763 | 3.0977154 | |
| Simulation | 744 | 120 | 2.9696395 | 2.31975072 | 3.8015975 | |
| Sports | 2256 | 50 | 0.4080622 | 0.29633429 | 0.5619152 | |
| Strategy | 473 | 188 | 7.3179953 | 5.82100004 | 9.199975 | |

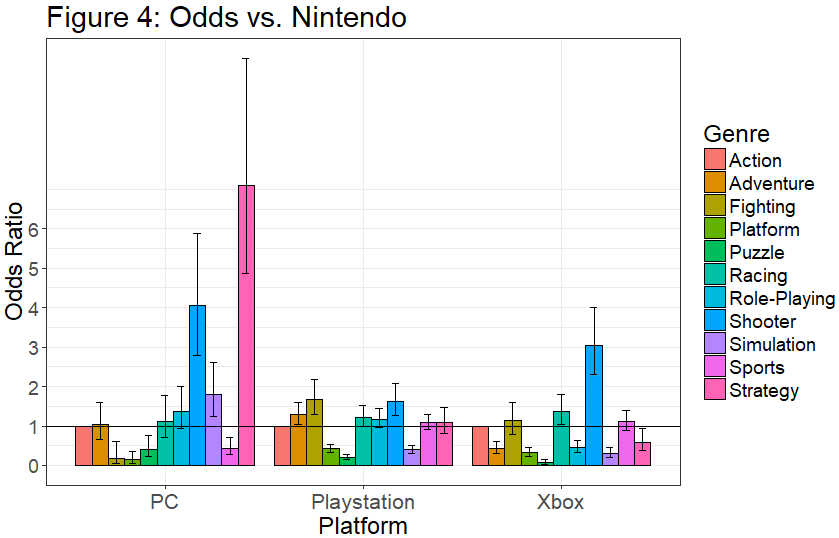
A similar association is found on the second grouping which separated the consoles into handheld and non-handheld consoles. The results showed that handhelds (Gameboy, PSP, etc.), are more than twice as likely to be a puzzle, adventure, or simulation game when compared to non-handheld consoles (Playstation, Xbox, etc). Shooter, sports and racing games are much less likely to be released on handhelds compared.

**Results for Genre Comparisons: Nintendo, Playstation, Xbox, and the PC**

The third type of grouping that was done categorized platforms based on their manufacturer. Titles released on the NES, SNES, N64, etc. are all included in Nintendo, titles released on the PS4, PS3, etc. are all grouped under Playstation, and titles released on the Xbox, Xbox 360, and Xbox One are categorized under Xbox. In summary, the four groups being compared are Playstation, Xbox, Nintendo, and the PC.

The hypothesis that there is no association between the genre of the game and which of the four platform types it is released on is rejected after conducting a chi-squared test for independence. The test statistic of 2083.5 with 30 degrees of freedom provides significant evidence indicating that genre is not independent of platform. Further chi-squared tests of each of the subsets (ie Consoles only, or a one on one comparison such as Nintendo vs. Playstation) all showed significant associations between genre and platform type.

To explore the associations, an analysis of odds ratios was done. A graphical display of the odds ratios can help to show which genres and platforms impact the association the most and is displayed in Figure 4. Since action games are the most popular genre for each of the platforms and have a similar proportion for each (making up 20 to 25% of the titles), they were used as the baseline genre for comparison. Nintendo was used as the baseline console. If a given 95% interval overlaps with 0, this indicates that there is not a significant difference in the odds of observing that genre when compared with Nintendo.

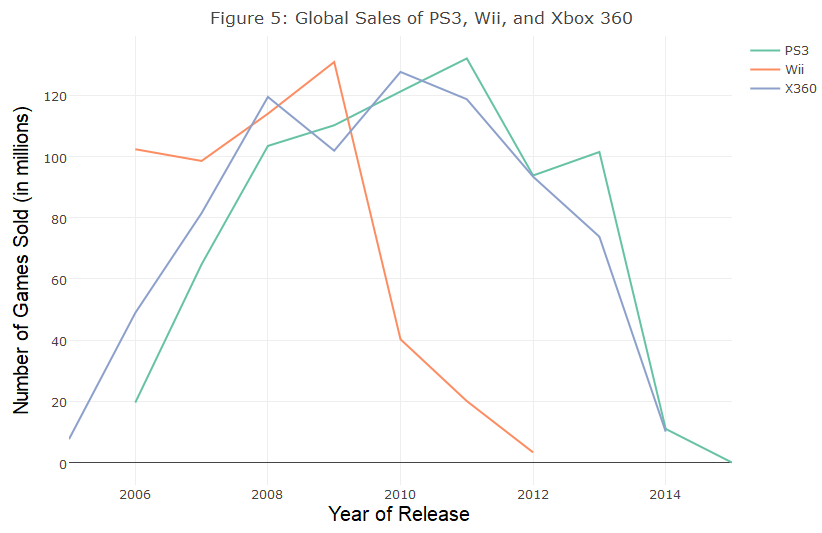


As before, the association between genre and platform can mainly be attributed to games released on the PC, with PC games being about 5-10 times more likely to have strategy games than Nintendo. There are two other main contributors to the association between genre and platform. The first is the decreased likelihood of platform and puzzle games on the PC, Playstation, and Xbox. Each of these estimations is far below the odds ratio of 1, which indicates that the Nintendo is significantly more likely (approximately 2 to 4 times more likely) to have platform and puzzle games than the others. The second is the increased likelihood of observing shooter games on the Xbox. Given that a title is going to be released on the Xbox, it is about three times more likely to be a shooter game than if it were being released on the Nintendo.

**Research Question 3: Are there any associations in sales or titles when we consider more current popular platforms?**

The previous test showed there was a significant association between the manufacturer and the genre of the game but when similar consoles are compared a different result is uncovered.

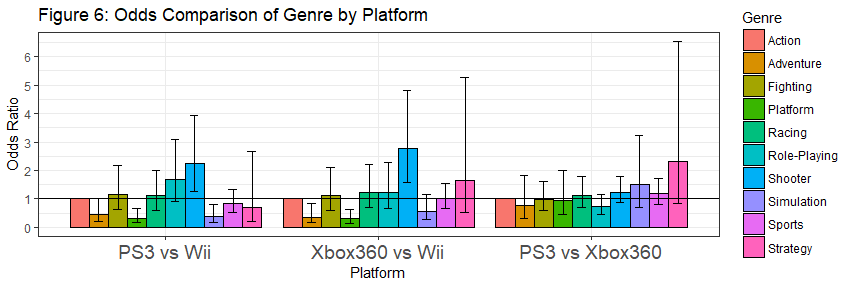
The Nintendo Wii was released in 2006, the Playstation 3 was released in 2006 and the Xbox 360 released in 2005. Not only are the three consoles released around the same time, but they also rank 2nd, 3rd, and 4th all time in sales. Based on the available data, the Xbox 360, Playstation 3, and Nintendo Wii have sold approximately 783 million, 757 million, and 509 million games respectively. These three platforms account for approximately 26% of the 8.7 billion video games units sold.



Sales trends over time show that the PS3 and Xbox360 are remarkably similar in their sales history. The Wii though differs in that its sales declined much sooner than the others. To explore the associations further, multiple chi-squared tests were done.

|  |  |
| --- | --- |
| **Table 4: Chi Squared Test Results Genre vs. Platform** | |
| **Platforms** | **Test Results** |
| Playstation3, Wii and Xbox360 | ꭓ2= 130.38, df = 20, p-value =0 |
| Playstation3 and Wii | ꭓ2= 80.753, df = 10, p-value = 0 |
| Wii and Xbox 360 | ꭓ2= 86.772, df = 10, p-value =0 |
| Playstation3 and Xbox 360 | **ꭓ2= 14.269, df = 10, p-value = 0.1611** |

The results displayed in Table 4 show significant evidence of an association between genre and platform, with the exception of the comparison directly between the PS3 and the Xbox 360. This test yields a p-value of 0.1611, which indicates that the distribution of titles is similar for these two platforms, with respect to genre. This leads to the suspicion that maybe only the Wii differs from the others.



Each of the intervals for the Ps3 vs. Xbox360 overlaps 0, indicating that there isn’t any increased likelihood of a game being released in a certain genre over another. This further supports the conclusion that the Playstation 3 and Xbox 360 tend towards the same types of games. Given that a title is being released on the Nintendo Wii, it is approximately two to five times more likely to be a platform game than if it were released on either the PS3 or the Xbox 360 but two to four times less likely to be a sports game.

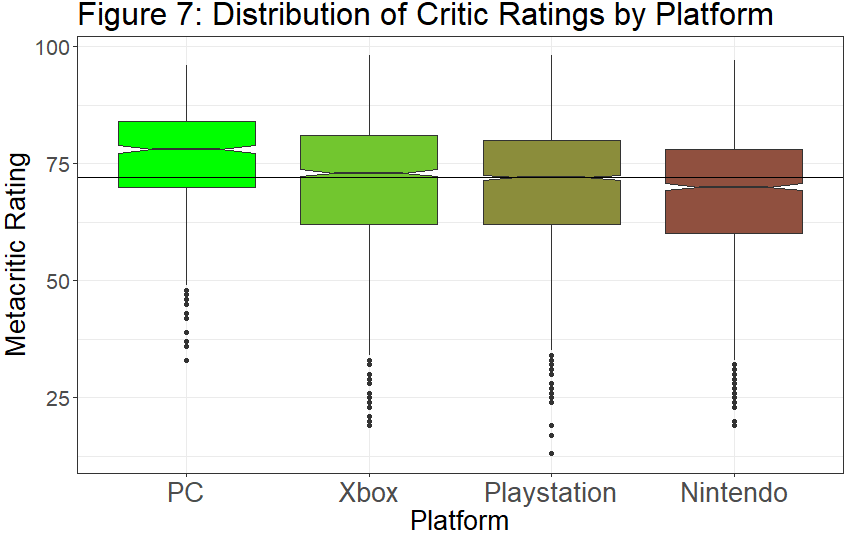
A further comparison of the more recent platforms, Playstation 4 and the Xbox One, shows that there is even less evidence of an association between these two platforms and genre (ꭓ2= 9.7345, df = 10, p-value =0.4641). Playstation and Xbox over time have become more similar with respect to the titles released.

**Research Question 4: Do critic ratings correlate with sales? Do critics tend to favor some publishers/genres/platforms over others?**

The website www.metacritic.com attempts to “average” critic ratings from a variety of sources. Metacritic scores are available for over 6,000 of the titles. Scores can range from 0 to 100 with 100 being the best rating possible. The overall median critic score was a 72, and scores ranged from 13 to 98. The highest rated game, Grand Theft Auto IV, had a score of 98 and sold 22.39 million copies worldwide. The highest selling game which was Wii Sports had a rating of 76.

An ANOVA for the critic scores from each manufacturer was computed and it was determined that at least one of the mean critic scores was significantly different than the others. Using Tukey’s method, a family of 95% confidence intervals was constructed and significantly different groups are shown in Table 5. Here again we observe that the only comparison that is not significantly different is that of Xbox vs. Playstation since the interval overlaps 0.

|  |  |  |
| --- | --- | --- |
| **Table 5: 95% Confidence intervals for the mean difference in Critic Rating.** | | |
| **Type** | **Lower Bound** | **Upper Bound** |
| PC-Nintendo | 6.6733168 | 9.924973 |
| Playstation-Nintendo | 1.2746409 | 3.47878 |
| Xbox-Nintendo | 1.3329644 | 3.843466 |
| Playstation-PC | -7.45347 | -.391399 |
| Xbox-PC | -.3556689 | -.066191 |
| Xbox-Playstation | -.9182772 | 1.341287 |



The most notable rating difference is that PC games are estimated to have critic score ratings between 6.67 to 9.9 points higher than Nintendo and roughly 4 to 7 points higher than the Xbox or Playstation on average. Nintendo games have the lowest mean ratings overall.

So do higher critic ratings tend towards higher sales? The distribution of the number of units sold is very sharply skewed to the right, ranging from 10,000 units sold to 78 million units sold. Since the distribution is so strongly skewed, a log transformation was done on global sales to make the distribution more symmetric.

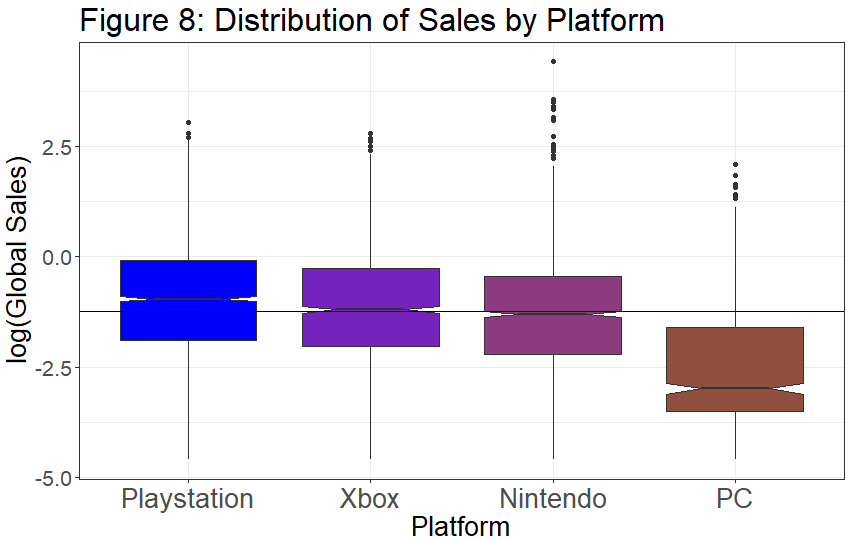
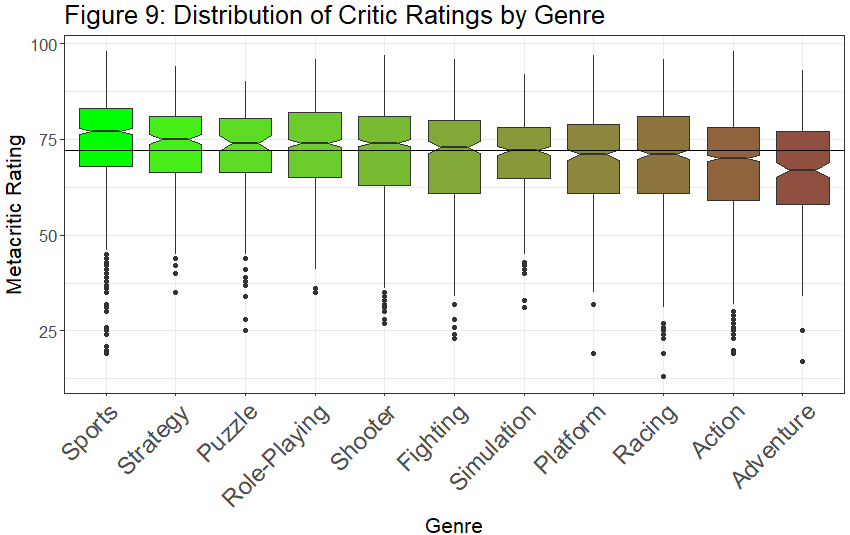


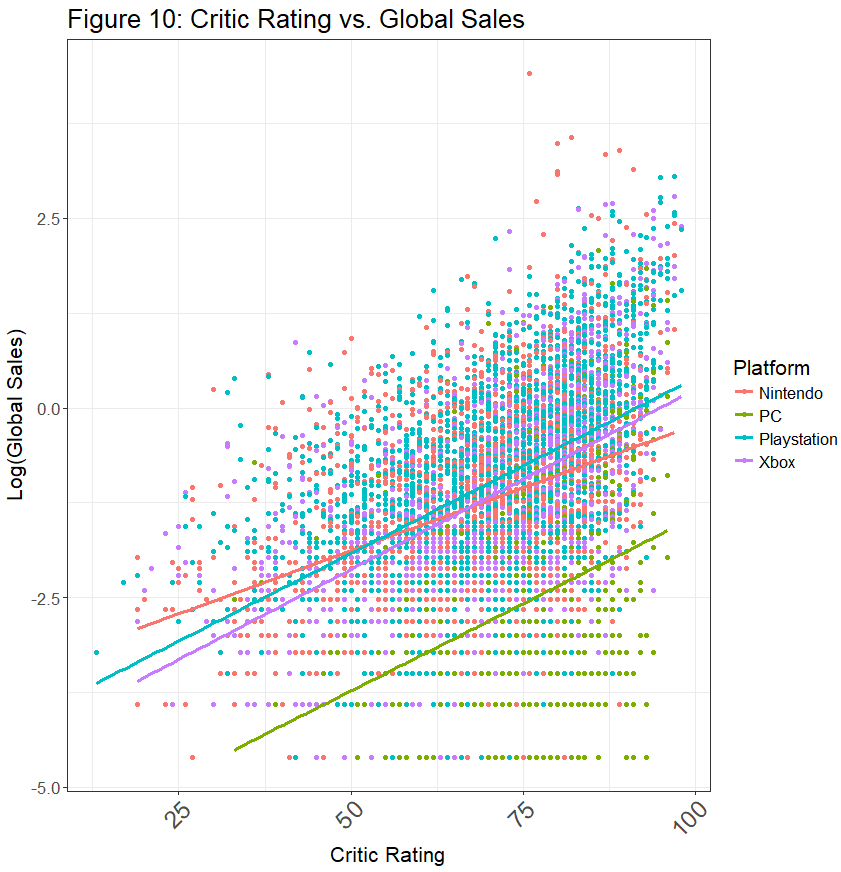
Figure 7 shows that games on the PC are generally rated higher than games on consoles, but as shown in Figure 8, fewer games are sold on the PC. This can partially be attributed to the affordability of a PC. Generally a computer system costs more than a console and so aren’t as likely to be purchased by the casual gamer.

It can also be shown that critic ratings are different by genre. An ANOVA with the hypothesis that the mean critic rating is equal for each genre yields an F Statistic of 19.4 (p-value<0.0001). This indicates that critics favor some genres over others. The distributions in Figure 9 reveal that sports games have the highest median rating and adventure games have the lowest.



After concluding that at least one of the ratings was significantly different, a family of confidence intervals was constructed using Tukey’s method. Significant results are displayed in Table 6. Sports games are the second best selling genre, so it makes sense that they would be ranked highly. Action games though sell more units globally than any other, but critic ratings are significantly lower than sports, strategy, role-playing and shooter games.

|  |  |  |
| --- | --- | --- |
| **Table 6: 95% Confidence Intervals**  **Mean Difference in Critic Rating by Genre** | | |
| **95% Intervals** | **Lower** | **Upper** |
| Sports-Racing | 2.338224179 | 6.9599508 |
| Sports-Shooter | 1.116972748 | 5.2409108 |
| Sports-Simulation | 1.301604711 | 7.1332808 |
| Sports-Adventure | 5.058001578 | 11.3244229 |
| Strategy-Adventure | 3.300992976 | 11.0356289 |
| Sports-Action | 4.526682971 | 8.1086419 |
| Strategy-Action | 2.405687481 | 8.1838348 |
| Role-Playing-Action | 2.935991867 | 6.875266 |
| Shooter-Action | 1.296522616 | 4.9809188 |

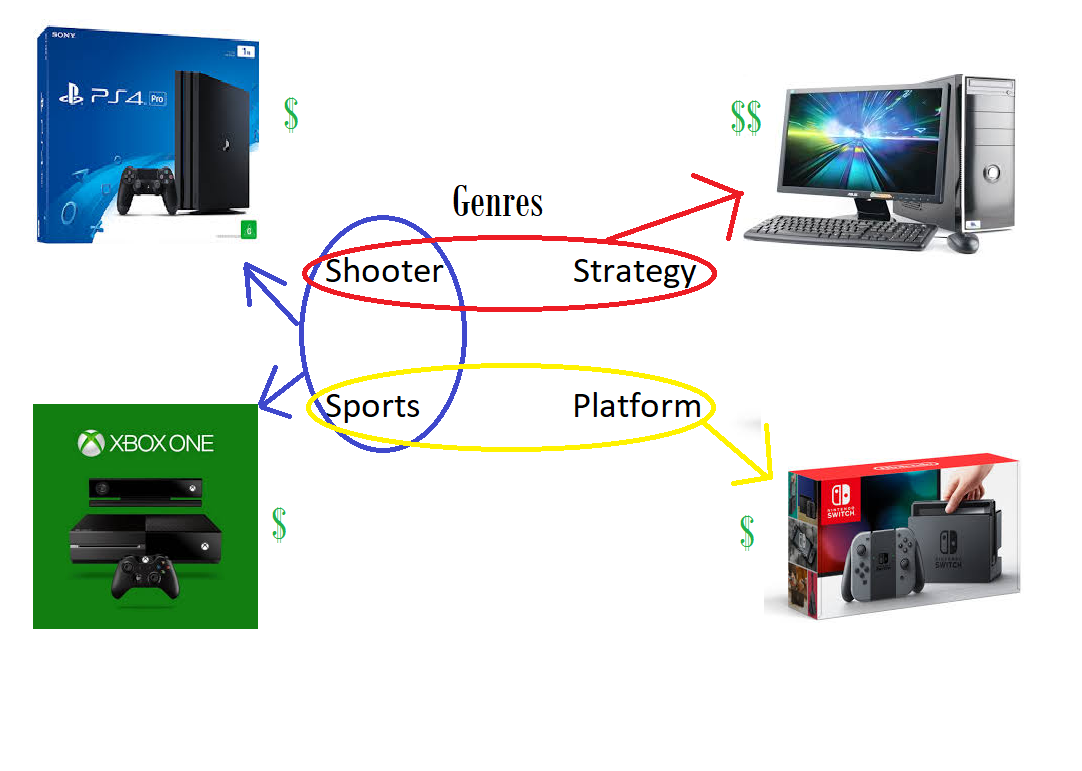
 Interestingly, strategy games have sold the least amount of games overall but have still are rated significantly higher. Although higher critic ratings for the PC do not mean that it sold the most, there is still generally a positive association between critic ratings and sales for each of the manufacturers.

|  |  |
| --- | --- |
| **Table 7: Correlation Between Rating and Sales** | |
| **Manufacturer** | **Correlation** |
| PC | 0.363455637 |
| Playstation | 0.495176736 |
| Xbox | 0.551996377 |
| Nintendo | 0.37067506 |

The estimated rate of sales increase for each additional point the critics rate a game is approximately the same for the PC, Playstation and Xbox when compared to the critic rating. Nintendo are not estimated to be as high for the most highly rated games and also not as low as Playstation and Xbox for the lower rated games. Table 7 shows that the correlation between sales and critic rating is similar for the Xbox and Playstation, but the association isn’t as strong for the PC and the Nintendo.

**Discussion of Findings**

Action games made up a similar portion of each platforms sales and titles. Other than this, the PC mostly had shooter, strategy, role-playing and simulation games. Playstation and Xbox showed similarities in that they and favored shooter and sports games. Nintendo showed differences from the Xbox and Playstation. Nintendo has had more successful platform games than the others. Most of the Mario Bros. games were platform games and these represent a large portion of the games that Nintendo has been successful with.



If a customer is looking to purchase a new system for video games, the types of games that they want to play will likely influence the decision. If they are looking for platform games and the ability to play sports and action games, the Nintendo may be the best option for them. If a customer is mainly looking for shooter games they could get a PC, Playstation, or Xbox. The difference is that if they also want to have strategy, simulation, and role playing games available, they should look into getting a PC. If they are more inclined to enjoy a sports game, they may be better off saving some money and getting an Xbox or Playstation rather than spending the extra money on a PC.

References

Agresti, Alan 2013, Categorical Data Analysis Third Edition

Chang, Winston 2013: *R Graphics Cookbook* *Practical Recipes for Visualizing Data*

DiChristopher, Tom CNBC Jan 2016 <https://www.cnbc.com/2016/01/26/digital-gaming-sales-hit-record-61-billion-in-2015-report.html>

Entertainment Software Association 2015 *Essential Facts About the Computer and Video Game Industry*, <http://www.theesa.com/wp-content/uploads/2015/04/ESA-Essential-Facts-2015.pdf>

# Ioana, Alex 2016 *Natural selection in home entertainment: on the best video games of all time*<https://deltanomics.wordpress.com/2016/12/27/natural-selection-in-home-entertainment-on-the-best-video-games-of-all-time/>

metacritic, 2017 CBS Interactive Inc., <http://www.metacritic.com/about-metacritic>

Video Game Charts, 2006-2017 VGChartz Ltd. <http://www.vgchartz.com/about.php>

Additional Data Source Information

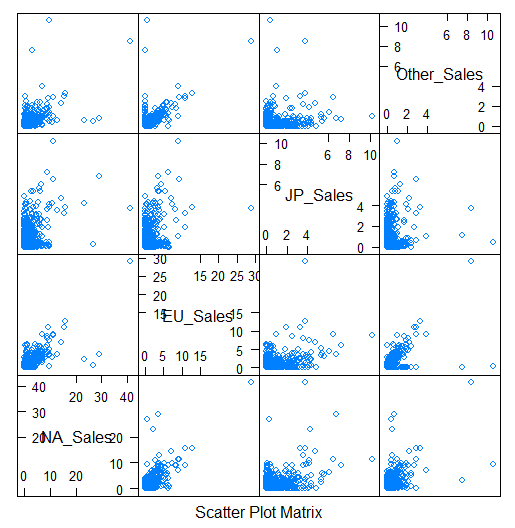
Video Game Charts (VGChartz.com), a business intelligence and research firm, publishes sales data for video games with a game database that includes over 40,000 titles.A set of 16,900 titles has been compiled by the Kaggle user Gregory Smith from the vgchartz.com website, and is published in public use data set section on Kaggle. The information gathered includes the publisher of the game, the platform that the game was released on, the year of release, and the genre of the game. Sales information is included from North America, Europe, Japan as well as total worldwide sales through the end of 2016. Only video game titles that had sales greater than 100,000 copies worldwide are included. Their methodology can be referenced on their website: <http://www.vgchartz.com/methodology.php>.

In addition to this information, the site metacritic.com gathers multiple ratings of video game titles and combines the ratings into a single “Metascore”for the game. This single score summarizes the many entertainment reviews that are available for the game. These critic scores are available for approximately 6900 of the video game titles in the vgchartz data and were added by Kaggle user Rush Kirubi.

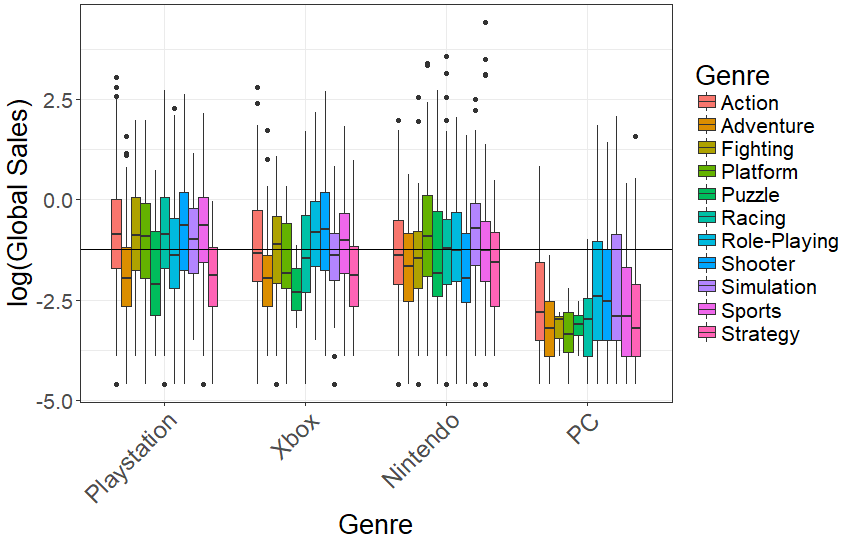
Data from vgchartz: <https://www.kaggle.com/gregorut/datasets>

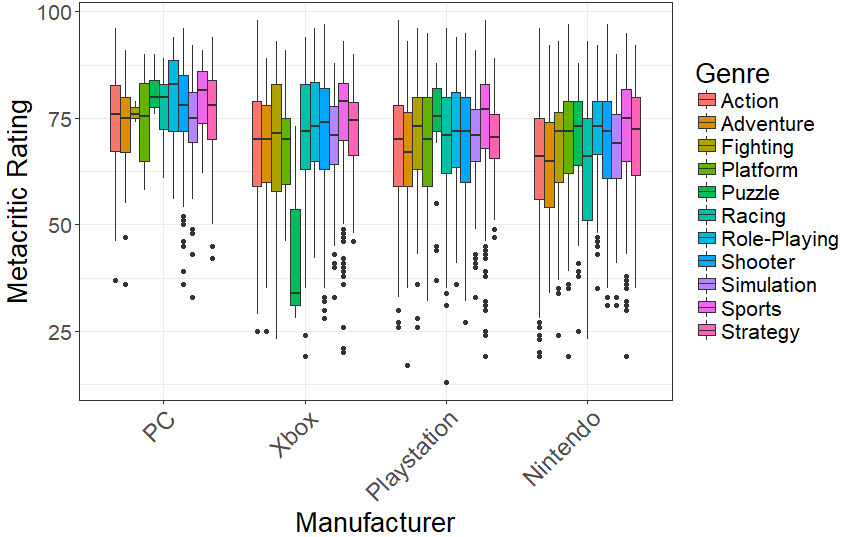
Sales data with added ratings: <https://www.kaggle.com/rush4ratio/video-game-sales-with-ratings>

Additional Charts and Tables



|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ***No. of Titles, Genre vs. Platform*** | | | | |
|  | **Nintendo** | **PC** | **PS** | **Xbox** |
| Action | 1166 | 170 | 1401 | 563 |
| Adventure | 434 | 65 | 673 | 88 |
| Fighting | 220 | 6 | 442 | 120 |
| Platform | 512 | 11 | 261 | 79 |
| Puzzle | 423 | 25 | 101 | 14 |
| Racing | 373 | 61 | 546 | 247 |
| Role-Playing | 520 | 104 | 734 | 112 |
| Shooter | 254 | 150 | 495 | 373 |
| Simulation | 457 | 120 | 219 | 68 |
| Sports | 792 | 50 | 1039 | 425 |
| Strategy | 182 | 188 | 239 | 52 |





R Code

###########  
#Author: Jacob Walsh  
#Title: Video Game Sales - Types and Trends  
  
library(MASS)  
library(tidyverse)  
library(stringr)  
library(magrittr)  
library(data.table)  
library(lubridate)  
library(RPostgreSQL)  
library(plotly)  
library(jsonlite)  
library(htmltools)  
library(glmnet)  
library(epitools)  
library(broom)  
library(lme4)  
library(sjPlot)  
library(parallel)  
library(pscl)  
library(countreg)  
  
  
vg.sales<-read.csv("C:/Penn State MAS/Stat 581/Video\_Games\_Sales\_as\_at\_22\_Dec\_2016.csv", na.strings=c("", "N/A", "Misc", "<NA>"))  
  
  
##############################New Categories and Variables#################################  
vg.sales$type1<-ifelse(vg.sales$Platform=="PC", "PC", "Console")   
vg.sales$type2<-ifelse(vg.sales$Platform=="PC", "PC",   
 ifelse(vg.sales$Platform %in% c("PSP", "PSV", "DS", "3DS", "GB"), "Handheld", "Console"))   
  
vg.sales$type3<-ifelse(vg.sales$Platform=="PC", "PC",   
 ifelse(vg.sales$Platform %in% c("Wii", "NES", "GB", "SNES", "WiiU", "3DS", "N64", "DS", "GBA", "GC"), "Nintendo",   
 ifelse(vg.sales$Platform %in% c("PS", "PS2", "PS3", "PS4", "PSP", "PSV"), "Playstation",   
 ifelse(vg.sales$Platform %in% c("XB", "X360", "XOne"), "Xbox", "NA"))))  
vg.sales$logsales<-log(vg.sales$Global\_Sales)  
  
vg.sales$scoresq<-vg.sales$Critic\_Score^2  
  
vg.sales<-vg.sales[!vg.sales$type3=="NA",]  
vg.sales<-vg.sales[!is.na(vg.sales$Genre),]  
vgc<-vg.sales[complete.cases(vg.sales),]  
  
GxT<-table(vg.sales$Genre, vg.sales$type3)  
totals<-aggregate(Global\_Sales~Name+Publisher, data=vg.sales, FUN=sum)  
sales=vg.sales[,c("NA\_Sales","EU\_Sales","JP\_Sales","Other\_Sales","Global\_Sales")]  
cor(sales)  
  
vg.sales$Platform<-factor(vg.sales$Platform)  
plat.table<-aggregate(vg.sales$Global\_Sales, by=list(vg.sales$Platform), FUN=sum)  
  
  
#######################Analysis of total sales  
##############Global Sales by Region  
global.trend<-aggregate(vgc$Global\_Sales, by=list(Year=vgc$Year\_of\_Release), FUN=sum)  
na.trend<-aggregate(vgc$NA\_Sales, by=list(Year=vgc$Year\_of\_Release), FUN=sum)  
eu.trend<-aggregate(vgc$EU\_Sales, by=list(Year=vgc$Year\_of\_Release), FUN=sum)  
jp.trend<-aggregate(vgc$JP\_Sales, by=list(Year=vgc$Year\_of\_Release), FUN=sum)  
other.trend<-aggregate(vgc$Other\_Sales, by=list(Year=vgc$Year\_of\_Release), FUN=sum)  
  
eu.sales<-eu.trend$x  
na.sales<-na.trend$x  
jp.sales<-jp.trend$x  
other.sales<-other.trend$x  
year<-na.trend$Year  
trend.df<-data.frame(year, na.sales, eu.sales, jp.sales, other.sales)  
  
vg.sales$Platform<-factor(vg.sales$Platform)  
plat.table<-aggregate(vg.sales$Global\_Sales, by=list(vg.sales$Platform), FUN=sum)  
  
#Console vs PC titles released  
  
t1<-table(vg.sales$Genre, vg.sales$type1)  
t1  
t1df<-as.data.frame(t1)  
prop.table(t1, 2)  
chisq.test(t1)  
tab<-epitab(t1)  
tabdf<-as.data.frame(tab$tab)  
tabdf[,c(1,3,5,6,7)]  
epitab(t1)  
  
t1df<-as.data.frame(prop.table(t1,2))  
colnames(t1df)<-c("Genre", "Platform", "Proportion")  
#Console vs PC vs Handheld  
  
t2<-table(vg.sales$Genre, vg.sales$type2)   
t2  
prop.table(t2, 2)  
chisq.test(t2)  
t2df<-as.data.frame(prop.table(t2,2))  
colnames(t2df)<-c("Genre", "Platform", "Proportion")  
epitab(t2[,c(1,2)])  
epitab(t2[,c(1,3)])  
  
t3<-table(vg.sales$Genre, vg.sales$type3)   
t3  
prop.table(t3, 2)  
chisq.test(t3)  
t3df<-as.data.frame(prop.table(t3,2))  
colnames(t3df)<-c("Genre", "Platform", "Proportion")  
p4<-ggplot(t3df, aes(x=Genre, y=Proportion, fill=Platform))+geom\_col(position="dodge", color="black")  
  
tab1<-epitab(t3[,c(1,2)], conf.level = 1-0.05/11)$tab  
tab2<-epitab(t3[,c(1,3)], conf.level = 1-0.05/11)$tab  
tab3<-epitab(t3[,c(1,4)], conf.level = 1-0.05/11)$tab  
odds<-c(tab1[,5:7], tab2[,5:7], tab3[,5:7])  
genres<-unique(factor(vgc$Genre))  
odds.df<-setDT(as.data.frame(matrix(odds, ncol=9, dimnames=list(rownames(tab1), c("PC", "PCmin", "PCmax", "Playstation", "PSmin", "PSmax", "Xbox", "Xmin", "Xmax")))), keep.rownames = TRUE)   
  
  
ggodds<-melt(odds.df, id.vars="rn", measure.vars=c("PC", "Playstation","Xbox"))   
ggmin<-melt(odds.df, id.vars="rn", measure.vars=c("PCmin", "PSmin","Xmin"))  
ggmax<-melt(odds.df, id.vars="rn", measure.vars=c("PCmax", "PSmax","Xmax"))  
ggodds$min<-ggmin$value  
ggodds$max<-ggmax$value  
colnames(ggodds)<-c("Genre", "Platform", "Odds", "min", "max")  
ggodds  
  
  
pxw<-vgc[vgc$Platform %in% c("PS3", "X360", "Wii"),]  
pxw$Platform<-factor(pxw$Platform)  
time.pxw<-aggregate(pxw$Global\_Sales, by=list(pxw$Year\_of\_Release, pxw$Platform), FUN=sum)  
f <- list(  
 family = "Arial",  
 size = 20,  
 color = "black"  
)  
a<-list(  
 family = "Arial",  
 size = 20,  
 color = "black"  
)  
  
pxw$Platform<-factor(pxw$Platform)  
t6<-table(pxw$Genre, pxw$Platform)  
  
chisq.test(t6)  
tab1<-epitab(t6[-5,c(2,1)], conf.level = 1-0.05/3)$tab  
tab2<-epitab(t6[-5,c(2,3)], conf.level = 1-0.05/3)$tab  
tab3<-epitab(t6[-5,c(1,3)], conf.level = 1-0.05/3)$tab  
  
odds<-c(tab1[,5:7], tab2[,5:7], tab3[,5:7])  
odds.df<-setDT(as.data.frame(matrix(odds, ncol=9, dimnames=list(rownames(t6[-5,]), c("PS3 vs Wii", "psmin", "psmax", "Xbox360 vs Wii", "xmin", "xmax", "PS3 vs Xbox360", "min", "max")))), keep.rownames = TRUE)   
  
  
ggodds2<-melt(odds.df, id.vars="rn", measure.vars=c("PS3 vs Wii", "Xbox360 vs Wii", "PS3 vs Xbox360"))   
ggmin<-melt(odds.df, id.vars="rn", measure.vars=c("psmin", "xmin","min"))  
ggmax<-melt(odds.df, id.vars="rn", measure.vars=c("psmax", "xmax","max"))  
ggodds2$min<-ggmin$value  
ggodds2$max<-ggmax$value  
colnames(ggodds)<-c("Genre", "Platform", "Odds", "min", "max")  
ggodds2  
  
######################Critic Score vs Sales############################  
  
anova=aov(vgc$Global\_Sales~vgc$type3)  
TukeyHSD((anova))  
  
lm1<-lm(logsales~Critic\_Score, data=vgc[vgc$type3=="PC",])  
summary(lm1)  
lm2<-lm(logsales~Critic\_Score, data=vgc[vgc$type3=="Playstation",])  
summary(lm2)  
lm3<-lm(logsales~Critic\_Score, data=vgc[vgc$type3=="Nintendo",])  
summary(lm3)  
lm4<-lm(logsales~Critic\_Score, data=vgc[vgc$type3=="Xbox",])  
summary(lm4)  
mod3<-glm(log(Global\_Sales)~Genre+Platform+Critic\_Score, data=vgc)  
summary(mod3)  
hist(mod3$residuals)  
  
tdf<-as.data.frame(table(pxw$Genre, pxw$Platform, pxw$Year\_of\_Release))  
colnames(tdf)<-c("Genre", "Console", "Year", "Freq")  
tdf$indicator<-ifelse(tdf$Freq==0, 1, 0)  
  
  
hom<-glm(Freq~Genre+Console+Year+Genre\*Console+Genre\*Year+Console\*Year+indicator, family=quasipoisson, data=tdf)  
cndlx<-glm(Freq~Genre+Console+Year+Genre\*Console+Genre\*Year+indicator, family=quasipoisson, data=tdf)  
cndly<-glm(Freq~Genre+Console+Year+Genre\*Console+Console\*Year+indicator, family=quasipoisson, data=tdf)  
cndlz<-glm(Freq~Genre+Console+Year+Console\*Year+Genre\*Year+indicator, family=quasipoisson, data=tdf)  
jntx<-glm(Freq~Genre+Console+Year+Console\*Year+indicator, family=quasipoisson, data=tdf)  
jnty<-glm(Freq~Genre+Console+Year+Genre\*Year+indicator, family=quasipoisson, data=tdf)  
jntz<-glm(Freq~Genre+Console+Year+Genre\*Console+indicator, family=quasipoisson, data=tdf)  
ind<-glm(Freq~Genre+Console+Year+indicator, family=quasipoisson, data=tdf)  
  
models<-matrix(round(c(hom$deviance,cndlx$deviance,cndly$deviance,cndlz$deviance,jntx$deviance,jnty$deviance,jntz$deviance,ind$deviance,   
 hom$df.residual,cndlx$df.residual,cndly$df.residual,cndlz$df.residual,jntx$df.residual,jnty$df.residual,jntz$df.residual,ind$df.residual), 3), ncol=2)  
  
rownames(models)<-c("(XY, XY, XZ)", "(XY, XZ)", "(XY, YZ)", "(XZ, YZ)","(X, YZ)","(Y,XZ)","(Z,XY)","(X,Y,Z)")  
colnames(models)<-c("G^2", "DF")  
models  
#################################  
 #Figures and Graphs#  
#################################  
  
c<-scales::seq\_gradient\_pal("blue")(seq(0,1,length.out=20))  
p1<-ggplot(plat.table, aes(x=reorder(Group.1, -x), y=x), fill=factor(Group.1))+  
 geom\_col(fill=c)+  
 theme\_bw()+  
 theme(title=(element\_text(size=12)),axis.title=element\_text(size=14), axis.text.x=element\_text(size=12, angle=90),axis.text.y=element\_text(size=12))+  
 labs(title="Figure 1: Global Sales by Platform", x="Platform", y="Global Sales (millions of units)")  
p1  
  
p2<-plot\_ly(trend.df, x=~year, y=~na.sales, name="North America", type='scatter', mode='lines+markers') %>%  
 add\_trace(y=~eu.sales, name="Europe", mode='lines+markers') %>%  
 add\_trace(y=~jp.sales, name='Japan', mode='lines+markers')%>%  
 add\_trace(y=~other.sales, name='Other', mode='lines+markers')%>%  
 layout(title="Figure 2: Total Video Game Sales by Year of Release", font=list(size=12,  
 color = 'black', face="bold"),xaxis=list(title="Year"), yaxis=list(title="Sales(in millions)"))  
p2  
  
  
p3<-ggplot(t1df, aes(x=Genre, y=Proportion, fill=Platform))+geom\_col(position="dodge", col="black")+  
 labs(title="Figure 3: Genres in Consoles vs. PC", x="Genre",y="Proportion")+  
 scale\_fill\_manual(values=c("#3406F4","thistle4"))+  
 theme\_bw()+  
 theme(text=(element\_text(size=18)), axis.text.x=element\_text(size=14,angle=45,hjust=1, vjust=1),plot.caption=element\_text(size=12))  
p3  
  
p4<-ggplot(ggodds, aes(x=Platform, y=Odds, fill=Genre))+  
 geom\_bar(position=position\_dodge(),stat="identity", col="black")+  
 labs(title="Figure 4: Odds vs. Nintendo", x="Platform", y="Odds Ratio")+  
 geom\_hline(yintercept=1)+  
 geom\_errorbar(aes(ymin=min, ymax=max), width=0.4, position=position\_dodge(.9))+  
 theme\_bw()+  
 theme(text=(element\_text(size=16)), axis.text.x=element\_text(size=16),plot.caption=element\_text(size=12))+  
 scale\_y\_continuous(breaks=seq(0,6,1))  
p4  
  
  
p5<-plot\_ly(time.pxw, x=~time.pxw$Group.1, y=~time.pxw$x, color=~time.pxw$Group.2)%>%  
 add\_lines() %>%  
   
 layout(title="Figure 5: Global Sales of PS3, Wii, and Xbox 360",xaxis=list(title="Year of Release", titlefont=a), yaxis=list(title="Number of Games Sold (in millions)", titlefont=a))  
p5  
  
p6<-ggplot(ggodds2, aes(x=Platform, y=Odds, fill=Genre, label=Platform))+  
 geom\_bar(position=position\_dodge(),stat="identity", col="black")+  
 labs(title="Figure 6: Odds Comparison of Genre by Platform",x="Platform", y="Odds Ratio")+  
 geom\_hline(yintercept=1)+  
 geom\_errorbar(aes(ymin=min, ymax=max), width=0.4, position=position\_dodge(.9))+  
 theme\_bw()+  
 theme(axis.text.x=element\_text(size=14), aspect.ratio=0.3)+  
 scale\_y\_continuous(breaks=seq(0,6,1))  
p6  
  
  
  
c<-scales::seq\_gradient\_pal("green")(seq(0,1,length.out=4))  
p7<- ggplot(vgc, aes(x=reorder(type3, -Critic\_Score, FUN=median), y=Critic\_Score)) +  
 geom\_boxplot(fill=c, notch=TRUE) +  
 labs(title="Figure 7: Distribution of Critic Ratings by Platform", y="Metacritic Rating", x="Platform") +  
 geom\_hline(yintercept=72)+  
 theme\_bw()+  
 theme(text=(element\_text(size=20)), axis.text.x=element\_text(size=20),plot.caption=element\_text(size=10))  
p7  
  
  
  
c<-scales::seq\_gradient\_pal("blue")(seq(0,1,length.out=4))  
p8<-ggplot(vgc, aes(x=reorder(type3, -logsales, FUN=median), y=logsales)) +  
 geom\_boxplot(fill=c, notch=TRUE) +  
 labs(title="Figure 8: Distribution of Sales by Platform",y="log(Global Sales)", x="Platform") +  
 geom\_hline(yintercept=-1.237874)+  
 theme\_bw()+  
 theme(text=(element\_text(size=20)), axis.text.x=element\_text(size=20),plot.caption=element\_text(size=10))  
p8  
  
c<-scales::seq\_gradient\_pal("green")(seq(0,1,length.out=11))  
  
p9<-ggplot(vgc, aes(x=reorder(Genre, -Critic\_Score, FUN=median), y=Critic\_Score)) +  
 geom\_boxplot(fill=c, notch=TRUE) +  
 labs(title="Figure 9: Distribution of Critic Ratings by Genre",y="Metacritic Rating", x="Genre") +  
 geom\_hline(yintercept=72)+  
 theme\_bw()+  
 theme(text=(element\_text(size=16)), axis.text.x=element\_text(size=18,angle=45,hjust=1, vjust=1),plot.caption=element\_text(size=14))  
p9  
  
p10<-ggplot(vgc, aes(x=Critic\_Score, y=logsales, color=type3))+  
 geom\_point()+  
 geom\_smooth(method="lm", se=FALSE, size=1.3)+  
 theme\_bw()+  
 theme(text=(element\_text(size=16)), axis.text.x=element\_text(size=18,angle=45,hjust=1, vjust=1),plot.caption=element\_text(size=14))+  
 labs(title="Figure 10: Critic Rating vs. Global Sales", x="Critic Rating", y="Log(Global Sales)")  
p10