What Video Game Sales Can Tell Us About "The Best Video Game"

Game"

Nicholas Branco
5 December 2017

1 Executive Summary

This paper focuses on examining video game sales data in order to find the best formula for a well-selling game. I looked at the sales of each of Nintendo's systems as Nintendo has a monopoly of the top fifteen slots in the data set and I looked at the genres of the top ten best-selling games in order to form my conclusion. The two graphs below summarize these findings.
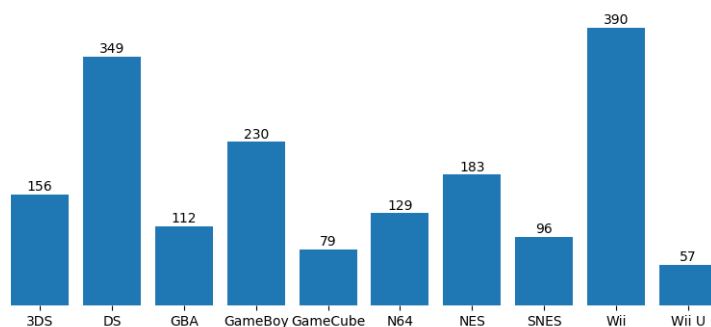

Figure 1.1: Nintendo's video game sales by system. The sales numbers are indicating the total global sales in millions for every video game in my data set on each system.
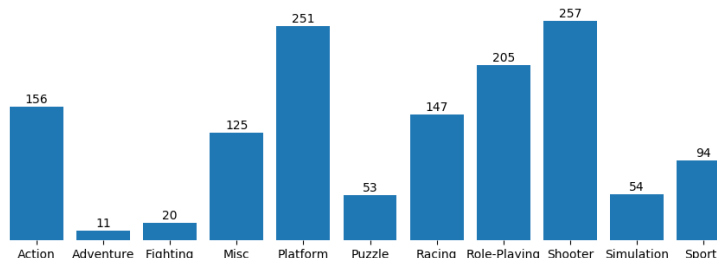

Figure 1.2: The sales of each genre that appear in the top 100, excluding Wii Sports. The sales numbers are indicating the total global sales in millions.

From this, I can conclude that a platforming video game or a shooter video game on a family-friendly system such as the Wii would be the optimal choice for creating the next best-selling video game. Since no system in the current generation of consoles (Wii U, Xbox One, and PlayStation 4) have the reputation of being family-friendly, then the best course of action would be to develop the game for the PlayStation 4 as it seems to be the most popular current generation console.

2 Introduction

What would you consider as the best video game ever made? I don't think that your choice will be the same as someone else's. For the consumers, what makes a good video game can be nearly anything. The scope of the world, the complexity of the story, the graphics, when the game was released, and when you played the game are all factors attributing to our idea of the best video game ever made. However, the consumers' idea of the best video game ever made is not as important as the developers'.

The video game industry is a business, and a business is out to make money. This changes the scope of our qualifications. We now have to take into account how well the video game did around the world. If a game sells well, then it will be made again (think every iteration of Call of Duty as an example) and improved upon. On the other hand, even if a great game sells poorly (take Metroid as an example - universally praised as great games but does not have the sales to back it up) the developer will move on to something new and forget about the old game for years.

This paper looks through a data set that lists an absurd amount of video games (about 16600) in order of global sales from most to least. It has some other information as well: genre, publisher, year released, platform, and sales for North America, Europe, Japan, and the rest of the world (grouped into an "other" column). This data set covers releases up until 2015 and a bit into 2016. Using this data set, I will be looking to answer the following questions in order to find what a video game developer can do to create a best-selling game.

3 Methodology

The first thing that had to be done was cleaning the data set. I did most of the cleaning in Linux's VI editor. There were many commas that were in the titles of video games that had to be removed or replaced, since the file was a .csv file. A few lines also had to be removed as they contained null values in important columns (chiefly, the global sales). They were extremely low on the list, so it didn't affect the results at all. I then checked that everything was okay using R to peek into the dataset. This code is available in appendix A.

The next thing was to find some basic statistics for the data set. I specifically wanted to find the most popular platform in the set. To do this, I wrote an R code to probe the data in a few ways. I looked at the current generations consoles (at the time of the data set they were the Wii U, Xbox One, and Playstation 4) and compared the sales. The Wii U was way below the other two (which I expected from my knowledge of the console however it contradicts the idea that Nintendo holds the top fifteen slots of the list). I then looked at the generation of consoles before that and found the Wii to be extremely popular. Finally, I compared the Wii's sales to the sales of the other systems that appear in the top 50 and I found that the Wii is still the most popular (I also checked the statistics without the Wii's number one seller for reasons

explained in the next paragraph and found that it still sat at the number one most popular console). This code is available in appendix B.

Looking at the data set, I found that the number one selling video game was Wii Sports. This wouldn't be an issue except for the fact that this game was not available to purchase by itself. It came packaged with the Wii and thus its sales are directly tied to the Wii's sales. This means that it was almost never bought because people liked that kind of game, but rather that it was bought only because consumers did not have a choice. I will specify when an observation is made excluding this record of data.

Once all of this was finished, I began the main code. Written in Python, the code takes all of the data and finds the following observations:

1. Nintendo's video game sales by platform.

2. Nintendo's video game sales by year.

3. Global sales of the top ten video games organized by genre.

4. Global sales of the top ten video games organized by genre, barring Wii Sports.

5. The frequency at which each of the genres of the top ten video games appear in the top 100 games.

6. The list of the top ten video games.

The reason I single out Nintendo for the first two observations is because of Nintendo's monopoly on the first fifteen games on this list. That impressive feat leads me to believe that they hold the key to making and selling great games. The code that creates the observations above is available in appendix C.

4 Data

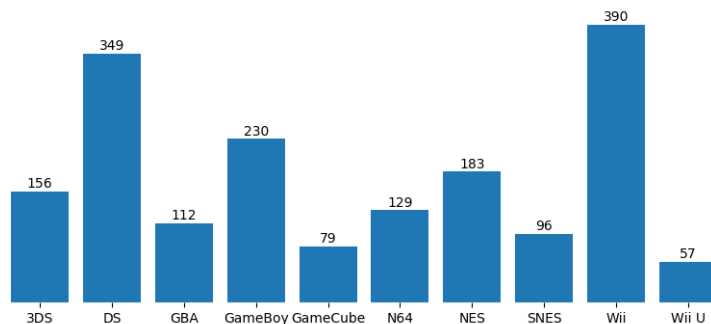First, I looked at Nintendo's video game sales by platform.

Figure 4.1: Nintendo's video game sales by system. The sales numbers are indicating the total global sales in millions for every video game in my data set on each system.

It is obvious that the Wii is the most popular console out of all of Nintendo's consoles. Indeed, from the R script that I had written, the Wii is definitely the most popular console, with or without the inclusion of Wii Sports. Interestingly, the Nintendo DS is a very close second place and will most certainly be included in the final analysis. As for the rest, there doesn't seem to be as much of a drastic change from one console to the next.
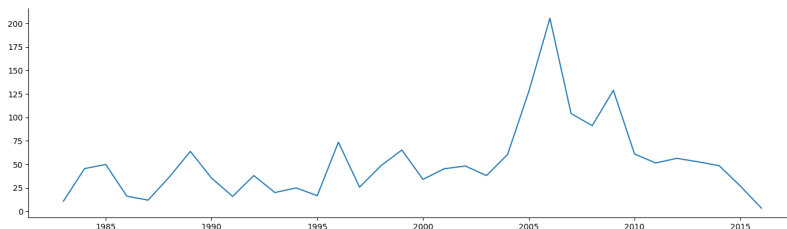
Next comes Nintendo's sales by year.

Figure 4.2: Nintendo's global sales in millions by year.

This was expected, since the DS and the Wii both released around 2005/2006 and so I expected peak sales around those years. This reinforces the idea that the Wii and DS are both systems that have games selling very well.

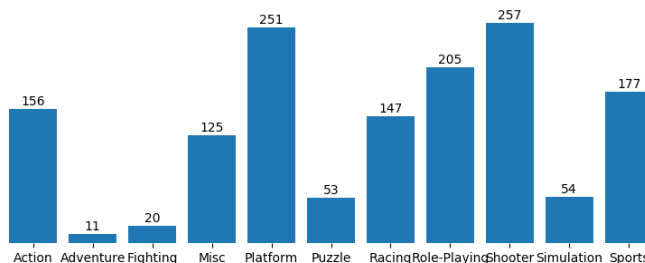For the next graph, I focus only on the top one hundred video games.

4

Figure 4.3: The sales of each genre that appear in the top 100. The sales numbers are indicating the total global sales in millions.

We can see that shooter games and platforming games are the major genres that sell well. These are still followed quite closely by role-playing games and to a lesser extent sports games. Knowing that sports games are not very popular with people who play video games (in general), I decided to create another graph that excludes Wii Sports to see how well the genre fares.
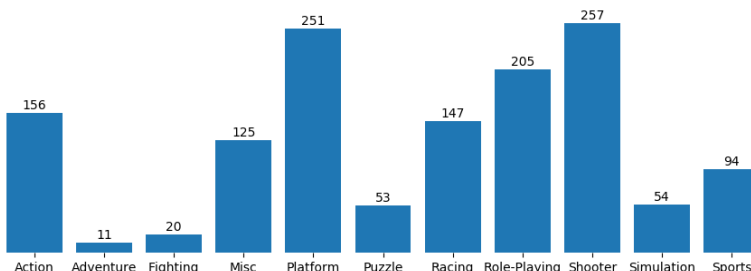
Figure 4.4: The sales of each genre that appear in the top 100, excluding Wii Sports. The sales numbers are indicating the total global sales in millions.

Obviously, Wii Sports is a great contributor to the sales of the sports genre of video games. This leaves shooting, platforming, and role-playing as the dominant genres of gaming. The extremely high sales of shooting games is very interesting as only the number 10 game on the list, Duck Hunt, is a shooter and the next is Call of Duty: Modern Warfare 3 at number 30. I speculate that the shooting genre selling better than the platforming genre is more due to an over saturation of shooting games rather than them selling very well (though they definitely sell very well and deserve a spot in the top few genres). This can be tested with one last observation.

Figure 4.5: The number of video games in the top 100 by genre. For this graph, Wii Sports was included.

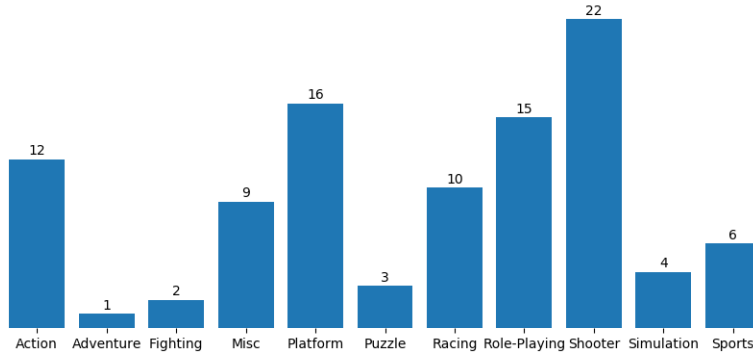While not much more, shooter games are much more common in the top 100 as platforming games, which may contribute to the genre selling so well. Every other genre seems about in the right spot in terms of the amount of video games in the genre and the total sales of the genre.

Finally, I want to list the top ten video games in the data set, as the exact nature of these games will be used to determine what kind of video game is the best bet to make the most money. The list is below.

Rank	Name	Platform	Year	Genre	Global Sales
1	Wii Sports	Wii	2006	Sports	82.74
2	Super Mario Bros.	NES	1985	Platform	40.24
3	Mario Kart Wii	Wii	2008	Racing	35.82
4	Wii Sports Resort	Wii	2009	Sports	33.00
5	Pokemon Red/Pokemon Blue	GB	1996	Role-Playing	31.37
6	Tetris	GB	1989	Puzzle	30.26
7	New Super Mario Bros.	DS	2006	Platform	30.01
8	Wii Play	Wii	2006	Misc	29.02
9	New Super Mario Bros. Wii	Wii	2009	Platform	28.62
10	Duck Hunt	NES	1984	Shooter	28.31

Note that Platforming games are the most popular in the top ten, while shooter games are limited to one entry: Duck Hunt. This list gives us an idea of what kind of games sell well. The Super Mario series of games seems to be the best-selling type of video game.

5   Conclusion

Looking at all of the data, I have found the following observations to be true:

1. The Wii is the best-selling console. It is closely followed by the Nintendo DS in terms of Nintendo's systems.

2. The shooting genre is the best-selling genre in the top 100 and is the most popular genre, having more video games in the top 100 than any other genre.

3. The platforming genre is very close in sales numbers to the shooting genre but with fewer games. This leads to the conclusion that platforming games generally sell better than shooting games, though both do very well in sales.

4. Platforming games are the most populous of the top 10, the three being Super Mario Bros. for the Nintendo Entertainment System, New Super Mario Bros. for the Nintendo DS, and New Super Mario Bros. Wii for the Wii.

All of these lead to the conclusion that the optimal type of game is a platforming video game, featuring a recognizable and well-loved character such as Mario, developed for a family-friendly system such as the Wii. The Wii was specifically chosen over the DS and the NES for its sheer sales numbers, being the most popular console. The Wii is unique as it was the first time that a video game system was found in more homes than not. It was popular with families with small children, elderly people, and adults alike. It was the perfect system to have in your home when hosting a party.

That being said, the Wii is an old system and, much like how the next generation of phones pretty much makes the previous versions obsolete, the newer generation of consoles makes the older generation obsolete. While most homes still have their Wii and play it from time to time, the majority of people have moved on to the next generation. This turns out to be a problem for our optimal game. The Wii U, the Wii's successor, has not sold well. This means that releasing a game on the Wii U is the worst way to reach the most people, unlike the Wii. This leaves us with two options: Xbox One and PlayStation 4.

While we can debate which of these systems is better, if we look at the sales of their video games, we can see that the PS4 is the best selling out of the current generation consoles (see Appendix B). This means that the best way to reach an audience is to develop the video game for the PS4. Most recent games are multi-console, though, so in time we could release for the Xbox One or even develop it for both systems at the same time. However, this analysis is assuming we are developing for only one system and the best one would be the PlayStation 4. This provides another problem.

Platforming games are very popular, but from the ones we've seen they have the added factor of featuring Mario, a recognizable character. Due to copyright, only Nintendo is able to use Mario in their video games. This leaves us with three options:

1. Create another character in the hopes that they become well-loved enough to sell games on their own.

2. Obtain the rights to an already-recognizable character and possibly pay a load of money in the process.

3. Switch genres.

Indeed, platforming games seem to be reliant, at least in part, by a recognizable character. As such, it may be impossible to acquire a character either through making one or paying money to use one. I think the third option would be best.

Switching the genre of the game we make is not too hard. Sure, the market is flooded with shooting games, but what is one more? Shooting games sell very well and reach a wide audience. That, coupled with the fact that shooter games are popular on both PS4 and Xbox One, leaves us with the final game: A shooter for the PS4.

This is only the beginning, however. Most shooter games benefit from multiplayer options, great graphics, a story that is just good enough to keep your attention (they are not known for their stories), and, if anything can be learned from Call of Duty, some form of Nazis. I merely provided the steps to arrive at what, sales-wise, is the best type of video game. Whether or not it will sell well will depend on the other factors mentioned. This study is limited in that it only could focus on the hard numbers and a familiarity with most of the video games on the list. This study was also limited with release dates. After this data was collected, Nintendo released another system (the Nintendo Switch) and at least two games that have made their way to become many people's favorite game (these would be The Legend of Zelda: Breath of the Wild and the more recent Super Mario Odyssey). I would love to further look at the sales numbers as more games are released for the Nintendo Switch, as it is the first console to merge both hand-held and home gaming, which changes nearly everything when it comes to accessibility. Now, Nintendo can not only release all of their games on one system (where they previously had two), those games can be played anywhere, meaning that the types of games that did better on hand-held consoles (such as turn-based RPG's like Pokemon or puzzle games like Tetris) may become more popular. In any case, the nature of the video game industry is constantly changing and one needs to be aware of the types of games that are popular at the time in order to maximize sales.

Appendix A: Cleaning the Data

This code is used to clean the data file vgsales.csv in order to get ready for analysis.

The data set has the following collumns:

Rank - Ranking of overall sales

Name - The games name

Platform - Platform of the game's release (i.e. PC,PS4, etc.)

Year - Year of the game's release

Genre - Genre of the game

Publisher - Publisher of the game

NA_Sales - Sales in North America (in millions)

EU_Sales - Sales in Europe (in millions)

JP_Sales - Sales in Japan (in millions)

Other_Sales - Sales in the rest of the world (in millions)

Global_Sales - Total worldwide sales.

There are a total of 16598 video games on this list. This list was obtained from https://www.kaggle.com/gregorut/videogamesales The data was changed so that all commas in video game titles were changed to spaces. This gives the right number of records. This was done using VI and removing all commas that were inbetween parenthesis. Then, I removed any other commas (like those in titles) on an individual basis by searching for lines with no values.

First, the imports. The tidyverse library is the only library we will be using in this code. We will be using the glimpse function to see how our data frames are arranged.

```r
library(tidyverse)
```

```
## Loading tidyverse: ggplot2
## Loading tidyverse: tibble
## Loading tidyverse: tidyr
## Loading tidyverse: readr
## Loading tidyverse: purrr
## Loading tidyverse: dplyr

## Conflicts with tidy packages ----------------------------------------------
## filter(): dplyr, stats
## lag():    dplyr, stats
```

Next, we read in the data. I have the original copy saved as "vgsales.csv" and the revised copy as "vgsales_commasremoved.csv".

```r
vg1 <- read.csv("vgsales_commasremoved.csv", header = TRUE, quote = "", fill = TRUE)
```

Let's check how the data looks.

```r
head(vg1)
```

```
##   Rank                   Name Platform Year     Genre Publisher
## 1    1             Wii Sports      Wii 2006    Sports  Nintendo
## 2    2      Super Mario Bros.      NES 1985  Platform  Nintendo
```

```
## 3     3            Mario Kart Wii     Wii 2008       Racing  Nintendo
## 4     4         Wii Sports Resort     Wii 2009       Sports  Nintendo
## 5     5 Pokemon Red/Pokemon Blue      GB 1996 Role-Playing  Nintendo
## 6     6                    Tetris      GB 1989       Puzzle  Nintendo
##   NA_Sales EU_Sales JP_Sales Other_Sales Global_Sales
## 1    41.49    29.02     3.77        8.46        82.74
## 2    29.08     3.58     6.81        0.77        40.24
## 3    15.85    12.88     3.79        3.31        35.82
## 4    15.75    11.01     3.28        2.96        33.00
## 5    11.27     8.89    10.22        1.00        31.37
## 6    23.20     2.26     4.22        0.58        30.26
```

tail(vg1)

```
##         Rank                                      Name Platform Year
## 16593 16595                                   Plushees       DS 2008
## 16594 16596      Woody Woodpecker in Crazy Castle 5        GBA 2002
## 16595 16597            Men in Black II: Alien Escape       GC 2003
## 16596 16598 SCORE International Baja 1000: The Official Game       PS2 2008
## 16597 16599                                Know How 2       DS 2010
## 16598 16600                         Spirits & Spells      GBA 2003
##            Genre  Publisher NA_Sales EU_Sales JP_Sales Other_Sales
## 16593 Simulation  Destineer     0.01     0.00        0           0
## 16594   Platform      Kemco     0.01     0.00        0           0
## 16595    Shooter  Infogrames     0.01     0.00        0           0
## 16596     Racing Activision     0.00     0.00        0           0
## 16597     Puzzle    7G//AMES     0.00     0.01        0           0
## 16598   Platform    Wanadoo     0.01     0.00        0           0
##       Global_Sales
## 16593         0.01
## 16594         0.01
## 16595         0.01
## 16596         0.01
## 16597         0.01
## 16598         0.01
```

Everything seems to be in order. Now, let's look at the header names to see what our data can tell us about.

names(vg1)

```
## [1] "Rank"        "Name"         "Platform"     "Year"
## [5] "Genre"       "Publisher"    "NA_Sales"     "EU_Sales"
## [9] "JP_Sales"    "Other_Sales"  "Global_Sales"
```

I will be using most of the columns in my analysis. The ones that I can specifically get rid of are EU_Sales and Other_Sales as I am focusing only on the global sales of each game. I am keeping in NA_Sales and JP_Sales as I may want to look at those later.

vg2 <- vg1[,**c**(1,2,3,4,5,6,7,9,11)]

And we check again.

head(vg2)

```
##   Rank                   Name Platform Year      Genre Publisher
## 1    1             Wii Sports      Wii 2006      Sports  Nintendo
## 2    2      Super Mario Bros.      NES 1985    Platform  Nintendo
## 3    3         Mario Kart Wii      Wii 2008      Racing  Nintendo
```

```
## 4    4           Wii Sports Resort      Wii 2009      Sports  Nintendo
## 5    5 Pokemon Red/Pokemon Blue     GB 1996 Role-Playing  Nintendo
## 6    6                     Tetris     GB 1989      Puzzle  Nintendo
##   NA_Sales JP_Sales Global_Sales
## 1    41.49     3.77        82.74
## 2    29.08     6.81        40.24
## 3    15.85     3.79        35.82
## 4    15.75     3.28        33.00
## 5    11.27    10.22        31.37
## 6    23.20     4.22        30.26
```

Good. Now,out of curiosity, I want to know what genres are listed. This will be important in determining the types of video games that sell well.

```
unique(vg2$Genre)
```

```
##  [1] Sports       Platform     Racing       Role-Playing Puzzle
##  [6] Misc         Shooter      Simulation   Action       Fighting
## [11] Adventure    Strategy
## 12 Levels: Action Adventure Fighting Misc Platform Puzzle ... Strategy
```

This is interesting. I expect platforming, role-playing, shooter, and adventure games to be the most popular, though from the top six games it seems like sports games may be giving them a run for their money.

The data is clean enough to begin analysis on, but I noticed that all of the platform names are abbreviated. I will expand their abbreviations so that if one comes up that is not well known, then it can be referenced here.

2600 : Atari 2600

3DO : 3DO Interactive Multiplayer (I have never heard of this before)

3DS : Nintendo 3DS

DC : Dreamcast

DS : Nintendo DS

GB : GameBoy/GameBoy Color

GBA : GameBoy Advance

GC : Gamecube

GEN : SEGA Genesis

GG : Game Gear

N64 : Nintendo 64

NES : Nintendo Entertainment System

NG : Neo Geo (Also have not heard of this one)

PC : Personal Computer

PCFX : PC-FX (Haven't heard of this one either)

PS : Playstation

PS2 : Playstation 2

PS3 : Playstation 3

PS4 : Playstation 4

PSP : Playstation Portable

PSV : Playstation Vita

SAT : SEGA Saturn

SCD : SEGA CD

SNES : Super Nintendo Entertainment System

TG16 : TurboGrafx-16

Wii : Nintendo Wii

WiiU : Nintendo Wii U

WS : WonderSwan

X360 : Xbox 360

XB : Xbox

XOne : Xbox One

Appendix B: Basic Data Statistics

This code is used to determine some basic statistics as they relate to my questions. We will again be working with the data "vgsales.csv". So let's begin by reading in the file.

```
vg <- read.csv("vgsales_commasremoved.csv", header = TRUE, quote = "", fill = TRUE)
```

To get an idea of what the data looks like, let's do this:

```
head(vg)
```

```
##   Rank                      Name Platform Year       Genre Publisher
## 1    1                Wii Sports      Wii 2006      Sports  Nintendo
## 2    2         Super Mario Bros.      NES 1985    Platform  Nintendo
## 3    3            Mario Kart Wii      Wii 2008      Racing  Nintendo
## 4    4         Wii Sports Resort      Wii 2009      Sports  Nintendo
## 5    5 Pokemon Red/Pokemon Blue       GB 1996 Role-Playing  Nintendo
## 6    6                    Tetris       GB 1989      Puzzle  Nintendo
##   NA_Sales EU_Sales JP_Sales Other_Sales Global_Sales
## 1    41.49    29.02     3.77        8.46        82.74
## 2    29.08     3.58     6.81        0.77        40.24
## 3    15.85    12.88     3.79        3.31        35.82
## 4    15.75    11.01     3.28        2.96        33.00
## 5    11.27     8.89    10.22        1.00        31.37
## 6    23.20     2.26     4.22        0.58        30.26
```

```
tail(vg)
```

```
##         Rank                                         Name Platform Year
## 16593 16595                                      Plushees       DS 2008
## 16594 16596         Woody Woodpecker in Crazy Castle 5      GBA 2002
## 16595 16597                Men in Black II: Alien Escape       GC 2003
## 16596 16598 SCORE International Baja 1000: The Official Game      PS2 2008
## 16597 16599                                   Know How 2       DS 2010
## 16598 16600                              Spirits & Spells      GBA 2003
##            Genre  Publisher NA_Sales EU_Sales JP_Sales Other_Sales
## 16593 Simulation  Destineer     0.01     0.00        0           0
## 16594   Platform      Kemco     0.01     0.00        0           0
## 16595    Shooter  Infogrames     0.01     0.00        0           0
## 16596     Racing  Activision     0.00     0.00        0           0
## 16597     Puzzle    7G//AMES     0.00     0.01        0           0
## 16598   Platform     Wanadoo     0.01     0.00        0           0
##       Global_Sales
## 16593         0.01
## 16594         0.01
## 16595         0.01
## 16596         0.01
## 16597         0.01
## 16598         0.01
```

Let's look at the current generation of video game consoles. I want to determine which is the most popular out of the three since most games are released on current generation consoles at the time of their release. These systems will be the Wii U, PlayStation 4, and Xbox One.

```r
wiiu <- vg[vg$Platform=="WiiU",]
ps4  <- vg[vg$Platform=="PS4",]
xb1  <- vg[vg$Platform=="XOne",]
```

Now, let's look at their total sales across the board:

```r
Reduce("+", wiiu$Global_Sales)
```

```
## [1] 81.86
```

```r
Reduce("+",ps4$Global_Sales)
```

```
## [1] 278.1
```

```r
Reduce("+",xb1$Global_Sales)
```

```
## [1] 141.06
```

From these, we can see that the PS4 has significantly more sales than the Wii U or the Xbox One. However, this may be due to the amount of records for the PS4 also being significantly higher than either of the other two. To check this, I am only going to take the first fifty records for each one.

```r
wiiu <- wiiu[1:50,]
ps4 <- ps4[1:50,]
xb1 <- xb1[1:50,]
```

And again, totaling the sales:

```r
Reduce("+",wiiu$Global_Sales)
```

```
## [1] 69
```

```r
Reduce("+",ps4$Global_Sales)
```

```
## [1] 189.32
```

```r
Reduce("+",xb1$Global_Sales)
```

```
## [1] 104.99
```

The result is still the same with the PS4 sitting on top, the Xbox One behind it, and the Wii U in dead last. Of course, these results seem correct since the Wii U was not known for having many amazing games on it and left the PS4 and Xbox One to be the more popular consoles of the generation, with the PS4 seeming to have surpassed the Xbox, at least in terms of the sales of the games.

This leads me to the conclusion that the more popular games (at least the newer ones) will be on the PS4 or Xbox One with a few Wii U games inbetween. To test this, let me check the top 10 recent games from 2016, the most recent year before this.

```r
recent <- vg[vg$Year=="2016",]
```

```r
head(recent,n=10)
```

```
##       Rank                     Name Platform Year      Genre
## 222    222                   FIFA 17      PS4 2016     Sports
## 272    272 Uncharted 4: A Thief's End      PS4 2016    Shooter
## 352    352  Tom Clancy's The Division      PS4 2016    Shooter
## 771    772            Far Cry: Primal      PS4 2016     Action
## 846    847  Tom Clancy's The Division     XOne 2016    Shooter
## 1027  1028                 Overwatch      PS4 2016    Shooter
## 1157  1158              No Man's Sky      PS4 2016     Action
```

```
## 1190 1191              Dark Souls III     PS4 2016 Role-Playing
## 1225 1226                     FIFA 17     XOne 2016       Sports
## 1390 1391                 Doom (2016)     PS4 2016      Shooter
##                         Publisher NA_Sales EU_Sales JP_Sales Other_Sales
## 222            Electronic Arts     0.28     3.75     0.06        0.69
## 272  Sony Computer Entertainment   1.30     2.07     0.18        0.65
## 352                    Ubisoft     1.28     1.61     0.15        0.57
## 771                    Ubisoft     0.59     1.16     0.06        0.33
## 846                    Ubisoft     1.20     0.62     0.00        0.18
## 1027                 Activision    0.64     0.68     0.14        0.26
## 1157               Hello Games    0.58     0.74     0.02        0.26
## 1190         Namco Bandai Games   0.58     0.44     0.33        0.21
## 1225           Electronic Arts    0.17     1.26     0.00        0.10
## 1390        Bethesda Softworks    0.49     0.66     0.02        0.22
##      Global_Sales
## 222          4.77
## 272          4.20
## 352          3.61
## 771          2.13
## 846          2.01
## 1027         1.73
## 1157         1.60
## 1190         1.56
## 1225         1.53
## 1390         1.39
```

As expected, most are PS4 games and only two are Xbox One games. There are no Wii U games on the list.

That's all well and good, but what console was the most popular? This will be important for finding out what type of console sells best (i.e. is it the more powerful or more gimmickey ones?)

I'm not going to bother with anything not in the top fifty platforms, as 50 is a good enough amount of titles to give the more popular consoles. These consoles will be:

```
top50 <- vg[1:50,]
unique(top50$Platform)
```

```
##  [1] Wii  NES  GB   DS   X360 PS3  PS2  SNES GBA  3DS  PS4  N64
## 31 Levels: 2600 3DO 3DS DC DS GB GBA GC GEN GG N64 NES NG PC PCFX ... XOne
```

So we have the Wii, NES, GameBoy, 3DS, DS, PlayStation 2, Super Nintendo, GameBoy Advance, and Nintendo 64. Lets take the top fifty entries from those so that we can compare with the current generation sales.

```
wii <- vg[vg$Platform=="Wii",]
nes  <- vg[vg$Platform=="NES",]
gb  <- vg[vg$Platform=="GB",]
ds3  <- vg[vg$Platform=="3DS",]
ds  <- vg[vg$Platform=="DS",]
ps2  <- vg[vg$Platform=="PS2",]
ps3  <- vg[vg$Platform=="PS3",]
snes  <- vg[vg$Platform=="SNES",]
gba  <- vg[vg$Platform=="GBA",]
n64  <- vg[vg$Platform=="N64",]
x360  <- vg[vg$Platform=="X360",]

wii <- wii[1:50,]
```

```
nes <- nes[1:50,]
gb <- gb[1:50,]
ds3 <- ds3[1:50,]
ds <- ds[1:50,]
ps2 <- ps2[1:50,]
ps3 <- ps3[1:50,]
snes <- snes[1:50,]
gba <- gba[1:50,]
n64 <- n64[1:50,]
xb360 <- x360[1:50,]
```

Okay, now let's list the values.

```
Reduce("+",nes$Global_Sales)
```

[1] 209.28

```
Reduce("+",gb$Global_Sales)
```

[1] 235.65

```
Reduce("+",ds3$Global_Sales)
```

[1] 164.3

```
Reduce("+",ds$Global_Sales)
```

[1] 341.51

```
Reduce("+",ps2$Global_Sales)
```

[1] 287.99

```
Reduce("+",snes$Global_Sales)
```

[1] 147.27

```
Reduce("+",gba$Global_Sales)
```

[1] 136.7

```
Reduce("+",n64$Global_Sales)
```

[1] 141.75

```
Reduce("+",wii$Global_Sales)
```

[1] 470.84

```
Reduce("+",ps3$Global_Sales)
```

[1] 315.82

```
Reduce("+",xb360$Global_Sales)
```

[1] 348.42

```
Reduce("+",ps4$Global_Sales)
```

[1] 189.32

So it turns out that Wii games are more popular. However, what happens when we ignore the Wii's top game Wii Sports? It was included with the system and so isn't too incicative of the sales of Wii games as a whole.

```r
wii <- vg[vg$Platform=="Wii",]
wii <- wii[2:51,]
Reduce("+",wii$Global_Sales)
```

```
## [1] 390.49
```

It's still the most popular.

So the Wii is the most popular system out of all of them, while the PS4 is the most popular current-generation console. The real analysis of the data and answering my data questions will begin in a Python script.

Appendix C: Python Code

```python
# Nicholas Branco
# 27 November 2017

# This code will take the video game sales data (cleaned-up
    ↪ version) and produce various graphs that will answer the
    ↪ following questions:

# 1. How have Nintendo's Sales increased over time with each
    ↪ system? (I chose Nintendo because they are my personal
    ↪ favorite and arguably the most popular publisher)
# 2. How have video game sales in general increased over each
    ↪ generation? (This one may not pan out)
# 3. Out of the top one hundred selling games, what are some
    ↪ common themes?
# 4. What do some of the bottom hundred games have in common?
    ↪ What can these tell us about which kinds of games sell
    ↪ better?

# So let's get started, shall we?

# Imports
import pandas as pd
import csv
import matplotlib.pyplot as plt
import numpy as np
import matplotlib.font_manager as font_manager


csvfile = open('vgsales.csv', 'rb')
reader = csv.reader(csvfile)
R = list(reader)

'''
Question One
How have video game sales for Nintendo changed over time? How
    ↪ have they changed
with each console? This part of the code generates graphs to
    ↪ visualize the data
and answer these questions.
'''

nint = [] # All games whose publisher is Nintendo
```

```
for row in R:
    if row[5] == 'Nintendo':
        nint.append(row)

'''
output = set()
for row in nint:
    output.add(row[2])
print output
'''
# From the above, I was able to find every system that Nintendo
    ↪ has published on: Nintendo 64, GameBoy Advance, 3DS, Wii,
    ↪ NES, GameCube, GameBoy, Wii U, Super Nintendo, and DS. Now
    ↪  I need to get each system into a dictionary with their
    ↪ global sales.

n64 = []
gba = []
ds3 = []
wii = []
nes = []
gc = []
gb = []
wiiu = []
snes = []
ds = []

for row in nint:
    if row[2] == 'N64':
        n64.append(float(row[10]))
    if row[2] == 'GBA':
        gba.append(float(row[10]))
    if row[2] == '3DS':
        ds3.append(float(row[10]))
    if row[2] == 'Wii':
        wii.append(float(row[10]))
    if row[2] == 'NES':
        nes.append(float(row[10]))
    if row[2] == 'GC':
        gc.append(float(row[10]))
    if row[2] == 'GB':
        gb.append(float(row[10]))
    if row[2] == 'WiiU':
        wiiu.append(float(row[10]))
    if row[2] == 'SNES':
        snes.append(float(row[10]))
```

```python
        if row[2] == 'DS':
            ds.append(float(row[10]))


systems = ["NES","GameBoy","SNES","N64","GBA","GameCube","DS","
    ↪ Wii","3DS","Wii␣U"]

sales = [sum(nes),sum(gb),sum(snes),sum(n64),sum(gba),sum(gc),sum
    ↪ (ds),sum(wii),
    sum(ds3),sum(wiiu)]

fig, ax = plt.subplots(figsize = (4,3))
ind = np.arange(len(systems))
rects1 = ax.bar(systems,sales)

def autolabel(rects):
    """
    Attach a text label above each bar displaying its height
    """
    for rect in rects:
        height = rect.get_height()
        ax.text(rect.get_x() + rect.get_width()/2., 1 + height,
                '%d' % float(height),
                ha='center', va='bottom')

autolabel(rects1)
ax.yaxis.set_visible(False)
ax.spines['top'].set_visible(False)
ax.spines['right'].set_visible(False)
ax.spines['bottom'].set_visible(False)
ax.spines['left'].set_visible(False)

plt.show()

# Of course, This tells us only which system has the most sales.
    ↪ How can we tell how their sales have increased over time?
    ↪ Let's take sales by year.

output = set()
for row in nint:
    output.add(row[3])

year_sales = dict()
x = 0
for year in output:
    x = 0
```

```python
        if year == 'N/A':
            q = 0
        else:
            for row in nint:
                if row[3] == year:
                    x = x + float(row[10])
                    year_sales[int(year)] = x

fig, ax = plt.subplots(figsize = (4,3))
#ax.scatter(year_sales.keys(),year_sales.values())
ax.plot(year_sales.keys(),year_sales.values())
#ax.yaxis.set_visible(False)
ax.spines['top'].set_visible(False)
ax.spines['right'].set_visible(False)
#ax.spines['bottom'].set_visible(False)
#ax.spines['left'].set_visible(False)
plt.show()



'''
Question Two
Out of the top 100 games, what do they have in common? To answer
    ↪ this, I will be
looking at publishers, genres, and platforms, and I will also be
    ↪ relying on my
personal knowledge of gaming in order to determine what, exactly,
    ↪  makes a good
game (by publisher's standards). This section of the code
    ↪ generates graphs that
show the number of video games that fall under each genre as well
    ↪  as each platform.
'''

output = set()
for row in R[1:101]:
    output.add(row[4])

genre_sales = dict()
x = 0
for genre in output:
    x = 0
    for row in R[1:101]:
        if row[4] == genre:
            x = x + float(row[10])
            genre_sales[genre] = x
```

```
fig, ax = plt.subplots(figsize = (4,3))
ind = np.arange(len(genre_sales.keys()))
rects1 = ax.bar(genre_sales.keys(),genre_sales.values())

def autolabel(rects):
    """
    Attach a text label above each bar displaying its height
    """
    for rect in rects:
        height = rect.get_height()
        ax.text(rect.get_x() + rect.get_width()/2., 1 + height,
                '%d' % float(height),
                ha='center', va='bottom')

autolabel(rects1)
ax.yaxis.set_visible(False)
ax.spines['top'].set_visible(False)
ax.spines['right'].set_visible(False)
ax.spines['bottom'].set_visible(False)
ax.spines['left'].set_visible(False)

plt.show()

# I decided to run this again without Wii Sports to see what
    ↪ would happen.

genre_sales = dict()
x = 0
for genre in output:
    x = 0
    for row in R[2:101]:
        if row[4] == genre:
            x = x + float(row[10])
            genre_sales[genre] = x

fig, ax = plt.subplots(figsize = (4,3))
ind = np.arange(len(genre_sales.keys()))
rects1 = ax.bar(genre_sales.keys(),genre_sales.values())

def autolabel(rects):
    """
    Attach a text label above each bar displaying its height
    """
    for rect in rects:
        height = rect.get_height()
```

```python
        ax.text(rect.get_x() + rect.get_width()/2., 1 + height,
                '%d' % float(height),
                ha='center', va='bottom')

autolabel(rects1)
ax.yaxis.set_visible(False)
ax.spines['top'].set_visible(False)
ax.spines['right'].set_visible(False)
ax.spines['bottom'].set_visible(False)
ax.spines['left'].set_visible(False)

plt.show()

# That is a MAJOR difference.

# That was sales, but now how about number. For this one, I won't
    ↪   have to exclude Wii Sports since it's sales have nothing
    ↪ to do with it.

genre_num = dict()
x = 0
for genre in output:
    x = 0
    for row in R[1:101]:
        if row[4] == genre:
            x = x + 1
            genre_num[genre] = x

fig, ax = plt.subplots(figsize = (4,3))
ind = np.arange(len(genre_num.keys()))
rects1 = ax.bar(genre_num.keys(),genre_num.values())

def autolabel(rects):
    """
    Attach a text label above each bar displaying its height
    """
    for rect in rects:
        height = rect.get_height()
        ax.text(rect.get_x() + rect.get_width()/2., 0.1 + height,
                '%d' % float(height),
                ha='center', va='bottom')

autolabel(rects1)
ax.yaxis.set_visible(False)
ax.spines['top'].set_visible(False)
ax.spines['right'].set_visible(False)
```

```
ax.spines['bottom'].set_visible(False)
ax.spines['left'].set_visible(False)

plt.show()

# Now, it's not super useful for me to be able to look at these
    ↪ graphs and not know what games I am talking about. So, I
    ↪ am going to print a simple list of the important
    ↪ information which will be pasted below so that it can be
    ↪ referenced in the report if need be.

for row in R[0:11]:
    print row[0:5]

'''
['Rank', 'Name', 'Platform', 'Year', 'Genre']
['1', 'Wii Sports', 'Wii', '2006', 'Sports']
['2', 'Super Mario Bros.', 'NES', '1985', 'Platform']
['3', 'Mario Kart Wii', 'Wii', '2008', 'Racing']
['4', 'Wii Sports Resort', 'Wii', '2009', 'Sports']
['5', 'Pokemon Red/Pokemon Blue', 'GB', '1996', 'Role-Playing']
['6', 'Tetris', 'GB', '1989', 'Puzzle']
['7', 'New Super Mario Bros.', 'DS', '2006', 'Platform']
['8', 'Wii Play', 'Wii', '2006', 'Misc']
['9', 'New Super Mario Bros. Wii', 'Wii', '2009', 'Platform']
['10', 'Duck Hunt', 'NES', '1984', 'Shooter']
'''
```