Video\_Games\_Case\_Study

Ralph Gracia

2023-11-10

## Setting Up My Environment

Before starting any work with the data, I made sure I had the necessary packages installed and ready to go.

install.packages("tidyverse")

## Installing package into '/cloud/lib/x86\_64-pc-linux-gnu-library/4.3'  
## (as 'lib' is unspecified)

library(tidyverse)

## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──  
## ✔ dplyr 1.1.3 ✔ readr 2.1.4  
## ✔ forcats 1.0.0 ✔ stringr 1.5.0  
## ✔ ggplot2 3.4.4 ✔ tibble 3.2.1  
## ✔ lubridate 1.9.3 ✔ tidyr 1.3.0  
## ✔ purrr 1.0.2

## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()  
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

install.packages("rmarkdown")

## Installing package into '/cloud/lib/x86\_64-pc-linux-gnu-library/4.3'  
## (as 'lib' is unspecified)

library(rmarkdown)

## Preparing The Data

I begun my work by taking a quick look at how the data is structured. Here we can get a better understanding of the variables and values stored within the data.

To view the source of my data click here [link](https://www.kaggle.com/datasets/ibriiee/video-games-sales-dataset-2022-updated-extra-feat)

video\_games <- read.csv("Video\_Games.csv")  
head(video\_games)

## Name Platform Year\_of\_Release Genre Publisher  
## 1 Wii Sports Wii 2006 Sports Nintendo  
## 2 Super Mario Bros. NES 1985 Platform Nintendo  
## 3 Mario Kart Wii Wii 2008 Racing Nintendo  
## 4 Wii Sports Resort Wii 2009 Sports Nintendo  
## 5 Pokemon Red/Pokemon Blue GB 1996 Role-Playing Nintendo  
## 6 Tetris GB 1989 Puzzle Nintendo  
## NA\_Sales EU\_Sales JP\_Sales Other\_Sales Global\_Sales Critic\_Score Critic\_Count  
## 1 41.36 28.96 3.77 8.45 82.53 76 51  
## 2 29.08 3.58 6.81 0.77 40.24 NA NA  
## 3 15.68 12.76 3.79 3.29 35.52 82 73  
## 4 15.61 10.93 3.28 2.95 32.77 80 73  
## 5 11.27 8.89 10.22 1.00 31.37 NA NA  
## 6 23.20 2.26 4.22 0.58 30.26 NA NA  
## User\_Score User\_Count Developer Rating  
## 1 8 322 Nintendo E  
## 2 NA   
## 3 8.3 709 Nintendo E  
## 4 8 192 Nintendo E  
## 5 NA   
## 6 NA

## Cleaning The Data

For this case study, my goal is to find information relevant to how games have performed from 2008 to early 2017 and what the most popular genres were. In order to achieve that, the first thing I needed to address was rows that had didn’t have a release year available in this dataset. I ultimately came to the conclusion that filtering out these rows wouldn’t affect the overall results.

last\_15\_years <- video\_games %>% filter(Year\_of\_Release >= 2008) %>%   
 arrange(-Global\_Sales)  
head(last\_15\_years)

## Name Platform Year\_of\_Release Genre  
## 1 Mario Kart Wii Wii 2008 Racing  
## 2 Wii Sports Resort Wii 2009 Sports  
## 3 New Super Mario Bros. Wii Wii 2009 Platform  
## 4 Kinect Adventures! X360 2010 Misc  
## 5 Wii Fit Plus Wii 2009 Sports  
## 6 Grand Theft Auto V PS3 2013 Action  
## Publisher NA\_Sales EU\_Sales JP\_Sales Other\_Sales Global\_Sales  
## 1 Nintendo 15.68 12.76 3.79 3.29 35.52  
## 2 Nintendo 15.61 10.93 3.28 2.95 32.77  
## 3 Nintendo 14.44 6.94 4.70 2.24 28.32  
## 4 Microsoft Game Studios 15.00 4.89 0.24 1.69 21.81  
## 5 Nintendo 9.01 8.49 2.53 1.77 21.79  
## 6 Take-Two Interactive 7.02 9.09 0.98 3.96 21.04  
## Critic\_Score Critic\_Count User\_Score User\_Count Developer Rating  
## 1 82 73 8.3 709 Nintendo E  
## 2 80 73 8 192 Nintendo E  
## 3 87 80 8.4 594 Nintendo E  
## 4 61 45 6.3 106 Good Science Studio E  
## 5 80 33 7.4 52 Nintendo E  
## 6 97 50 8.2 3994 Rockstar North M

The next step was to create new data frames for each console and see how well each console’s games have performed in each region since from 2008-2017.

#### Playstation

ps\_games <- last\_15\_years %>% filter(Platform =="PS" | Platform =="PS2" | Platform =="PS3" | Platform =="PSV" | Platform =="PSP"| Platform =="PS4") %>%  
 filter(!is.na(Year\_of\_Release)) %>% arrange(-Global\_Sales)  
head(ps\_games)

## Name Platform Year\_of\_Release Genre  
## 1 Grand Theft Auto V PS3 2013 Action  
## 2 Call of Duty: Black Ops 3 PS4 2015 Shooter  
## 3 Call of Duty: Black Ops II PS3 2012 Shooter  
## 4 Call of Duty: Modern Warfare 3 PS3 2011 Shooter  
## 5 Call of Duty: Black Ops PS3 2010 Shooter  
## 6 Grand Theft Auto V PS4 2014 Action  
## Publisher NA\_Sales EU\_Sales JP\_Sales Other\_Sales Global\_Sales  
## 1 Take-Two Interactive 7.02 9.09 0.98 3.96 21.04  
## 2 Activision 6.03 5.86 0.36 2.38 14.63  
## 3 Activision 4.99 5.73 0.65 2.42 13.79  
## 4 Activision 5.54 5.73 0.49 1.57 13.32  
## 5 Activision 5.99 4.37 0.48 1.79 12.63  
## 6 Take-Two Interactive 3.96 6.31 0.38 1.97 12.61  
## Critic\_Score Critic\_Count User\_Score User\_Count  
## 1 97 50 8.2 3994  
## 2 NA NA NA  
## 3 83 21 5.3 922  
## 4 88 39 3.2 5234  
## 5 88 58 6.4 1094  
## 6 97 66 8.3 2899  
## Developer Rating  
## 1 Rockstar North M  
## 2   
## 3 Treyarch M  
## 4 Infinity Ward, Sledgehammer Games M  
## 5 Treyarch M  
## 6 Rockstar North M

#### Xbox

xbox\_games <- last\_15\_years %>% filter(Platform == "XB" | Platform == "X360" | Platform == "XOne") %>%  
 filter(!is.na(Year\_of\_Release)) %>% arrange(-Global\_Sales)  
head(xbox\_games)

## Name Platform Year\_of\_Release Genre  
## 1 Kinect Adventures! X360 2010 Misc  
## 2 Grand Theft Auto V X360 2013 Action  
## 3 Call of Duty: Modern Warfare 3 X360 2011 Shooter  
## 4 Call of Duty: Black Ops X360 2010 Shooter  
## 5 Call of Duty: Black Ops II X360 2012 Shooter  
## 6 Call of Duty: Modern Warfare 2 X360 2009 Shooter  
## Publisher NA\_Sales EU\_Sales JP\_Sales Other\_Sales Global\_Sales  
## 1 Microsoft Game Studios 15.00 4.89 0.24 1.69 21.81  
## 2 Take-Two Interactive 9.66 5.14 0.06 1.41 16.27  
## 3 Activision 9.04 4.24 0.13 1.32 14.73  
## 4 Activision 9.70 3.68 0.11 1.13 14.61  
## 5 Activision 8.25 4.24 0.07 1.12 13.67  
## 6 Activision 8.52 3.59 0.08 1.28 13.47  
## Critic\_Score Critic\_Count User\_Score User\_Count  
## 1 61 45 6.3 106  
## 2 97 58 8.1 3711  
## 3 88 81 3.4 8713  
## 4 87 89 6.3 1454  
## 5 83 73 4.8 2256  
## 6 94 100 6.3 2698  
## Developer Rating  
## 1 Good Science Studio E  
## 2 Rockstar North M  
## 3 Infinity Ward, Sledgehammer Games M  
## 4 Treyarch M  
## 5 Treyarch M  
## 6 Infinity Ward M

#### Nintendo

nintendo\_games <- last\_15\_years %>% filter(Platform =="DS" | Platform =="3DS" | Platform =="Wii" | Platform =="WiiU") %>%  
 filter(!is.na(Year\_of\_Release)) %>% arrange(-Global\_Sales)  
head(nintendo\_games)

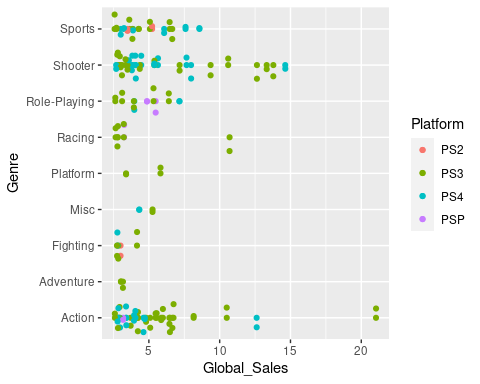
## Name Platform Year\_of\_Release Genre Publisher  
## 1 Mario Kart Wii Wii 2008 Racing Nintendo  
## 2 Wii Sports Resort Wii 2009 Sports Nintendo  
## 3 New Super Mario Bros. Wii Wii 2009 Platform Nintendo  
## 4 Wii Fit Plus Wii 2009 Sports Nintendo  
## 5 Pokemon Black/Pokemon White DS 2010 Role-Playing Nintendo  
## 6 Pokemon X/Pokemon Y 3DS 2013 Role-Playing Nintendo  
## NA\_Sales EU\_Sales JP\_Sales Other\_Sales Global\_Sales Critic\_Score Critic\_Count  
## 1 15.68 12.76 3.79 3.29 35.52 82 73  
## 2 15.61 10.93 3.28 2.95 32.77 80 73  
## 3 14.44 6.94 4.70 2.24 28.32 87 80  
## 4 9.01 8.49 2.53 1.77 21.79 80 33  
## 5 5.51 3.17 5.65 0.80 15.14 NA NA  
## 6 5.28 4.19 4.35 0.78 14.60 NA NA  
## User\_Score User\_Count Developer Rating  
## 1 8.3 709 Nintendo E  
## 2 8 192 Nintendo E  
## 3 8.4 594 Nintendo E  
## 4 7.4 52 Nintendo E  
## 5 NA   
## 6 NA

## Visualizing Trends

My original objective is to review the data and learn what the top genres of games were on each console based on their top 100 games. After analyzing the new data frames, I used ggplot to visualize my findings.

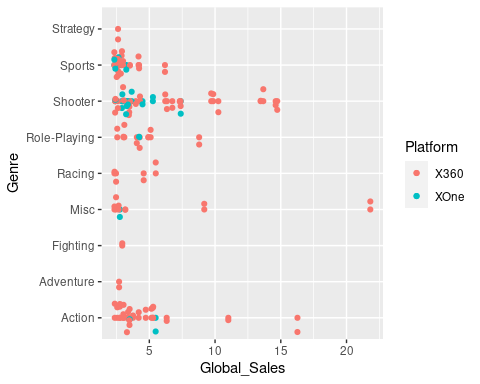
### Top Playstation Genres (Million USD)

ggplot(ps\_games %>% filter(Global\_Sales >= 2.58), aes(Global\_Sales, Genre, col=Platform)) + geom\_point() + geom\_jitter()



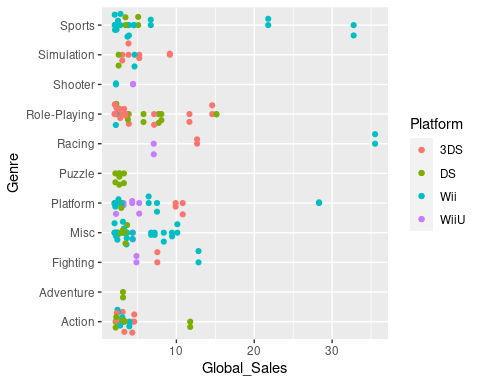
### Top Xbox Genres (Million USD)

ggplot(xbox\_games %>% filter(Global\_Sales >= 2.34), aes(Global\_Sales, Genre, col=Platform)) + geom\_point() + geom\_jitter()



### Top Nintendo Genres (Million USD)

ggplot(nintendo\_games %>% filter(Global\_Sales >= 2.06), aes(Global\_Sales, Genre, col=Platform)) + geom\_point() + geom\_jitter()



## Key Takeaways

As shown in the data visuals, the most popular genres on Playstation were:

* Action
* Shooter
* Sports

On Xbox:

* Action
* Shooter
* Role-Playing
* Sports

On Nintendo:

* Action
* Miscellaneous
* Platform
* Role-Playing

The following table further supports this discovery by displaying the top 50 games across all consoles.

top\_console\_games <- last\_15\_years %>% filter(Global\_Sales >= 7.38) %>%  
 arrange(-Global\_Sales)  
head(top\_console\_games)

## Name Platform Year\_of\_Release Genre  
## 1 Mario Kart Wii Wii 2008 Racing  
## 2 Wii Sports Resort Wii 2009 Sports  
## 3 New Super Mario Bros. Wii Wii 2009 Platform  
## 4 Kinect Adventures! X360 2010 Misc  
## 5 Wii Fit Plus Wii 2009 Sports  
## 6 Grand Theft Auto V PS3 2013 Action  
## Publisher NA\_Sales EU\_Sales JP\_Sales Other\_Sales Global\_Sales  
## 1 Nintendo 15.68 12.76 3.79 3.29 35.52  
## 2 Nintendo 15.61 10.93 3.28 2.95 32.77  
## 3 Nintendo 14.44 6.94 4.70 2.24 28.32  
## 4 Microsoft Game Studios 15.00 4.89 0.24 1.69 21.81  
## 5 Nintendo 9.01 8.49 2.53 1.77 21.79  
## 6 Take-Two Interactive 7.02 9.09 0.98 3.96 21.04  
## Critic\_Score Critic\_Count User\_Score User\_Count Developer Rating  
## 1 82 73 8.3 709 Nintendo E  
## 2 80 73 8 192 Nintendo E  
## 3 87 80 8.4 594 Nintendo E  
## 4 61 45 6.3 106 Good Science Studio E  
## 5 80 33 7.4 52 Nintendo E  
## 6 97 50 8.2 3994 Rockstar North M

## The Next Step?

* Incorporate past success to future inspirations
* Learn from indie developers to evaluate popularity

My suggestion to these three companies is to take a step back and recall the formulas that allowed earlier games to succeed to begin with as well as utilize indie developers to get a better understanding of new preferences among players.