

The background of the slide features a low-angle, upward-looking perspective of several modern skyscrapers. The buildings are dark in color, and their repetitive window patterns create a strong sense of verticality and geometric structure. The sky is a uniform, dark grey, which makes the white and yellow text stand out prominently.

# Final Project

---

By Hung Wen Chen

---

10 / 27 / 2022

# Video Games Sales 2019

# CONTENT

01

Data Review

02

New Attributes

03

Observation

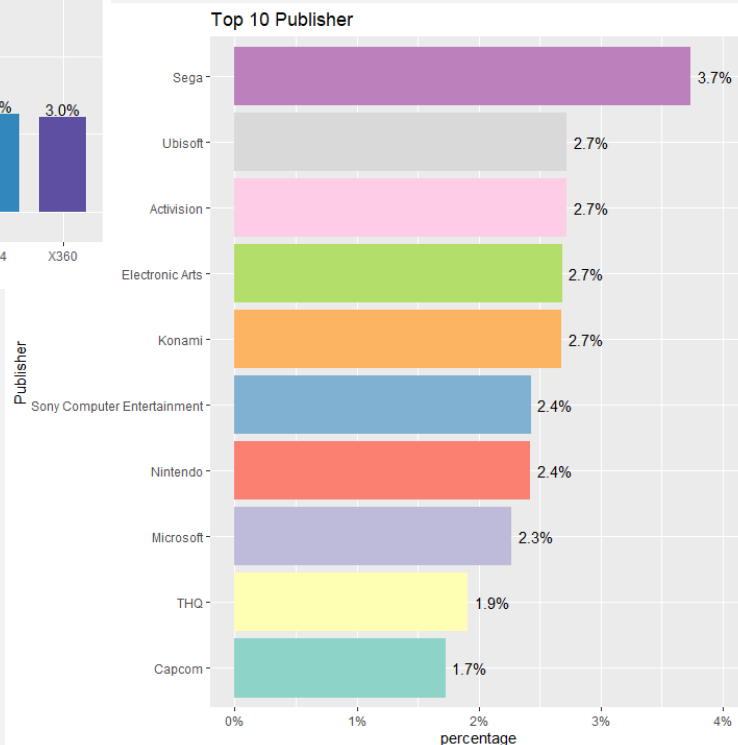
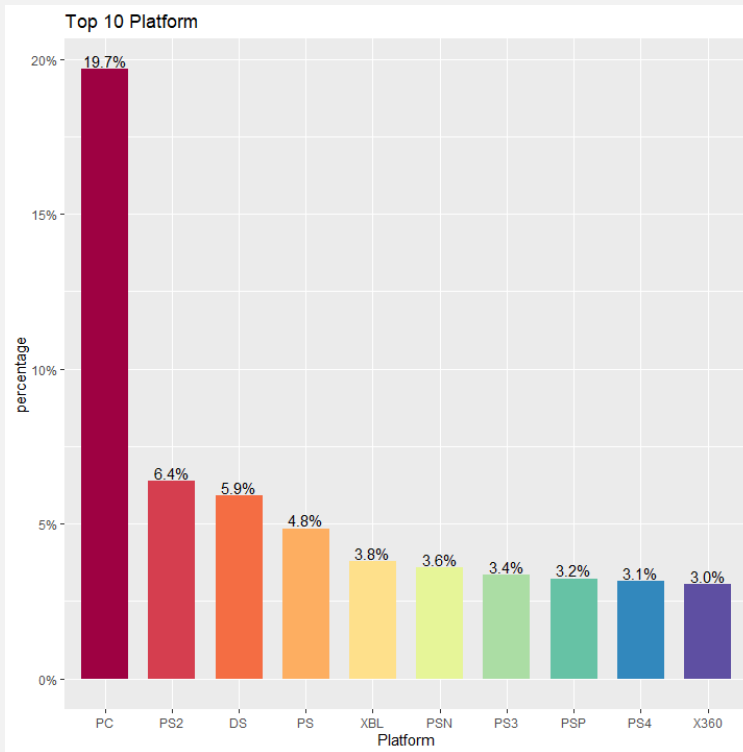
# I: Dataset Overview

---

Rank	Int
Name	Chr
Genre	Chr
ESRB_Rating	Chr
Platform	Chr
Publisher	Chr
Developer	Chr
Critic_Score	Numeric
User_Score	Numeric
Total_Shipped	Numeric
Global_Sales	Numeric
NA_Sales	Numeric
PAL_Sales	Numeric
JP_Sales	Numeric
Other_Sales	Numeric
Year	Numeric

- The dataset “vgsales-12-4-2019-short.csv” has 55,792 rows (some of them were same but different **Platform**) and 17 attributes.
- The original dataset (Rank) ordered by the **Total\_Shipped** and **Global\_Sales**.
- If the games have **Total\_Shipped**, it don't have Regional Sales data (Global, NA, PAL...).
- Many Publisher were **Unknown**.

# I: Dataset Overview



- There're 74 platforms and 3069 publishers in this dataset, which is the majority of them? These two graph give the explanation.
- Many publisher were **Unknown** (8.8%), so I avoid those data in the graph in order to see the true rank of them.
- We can see most of the games were mainly on the PC platform, but Publisher were distributed evenly.

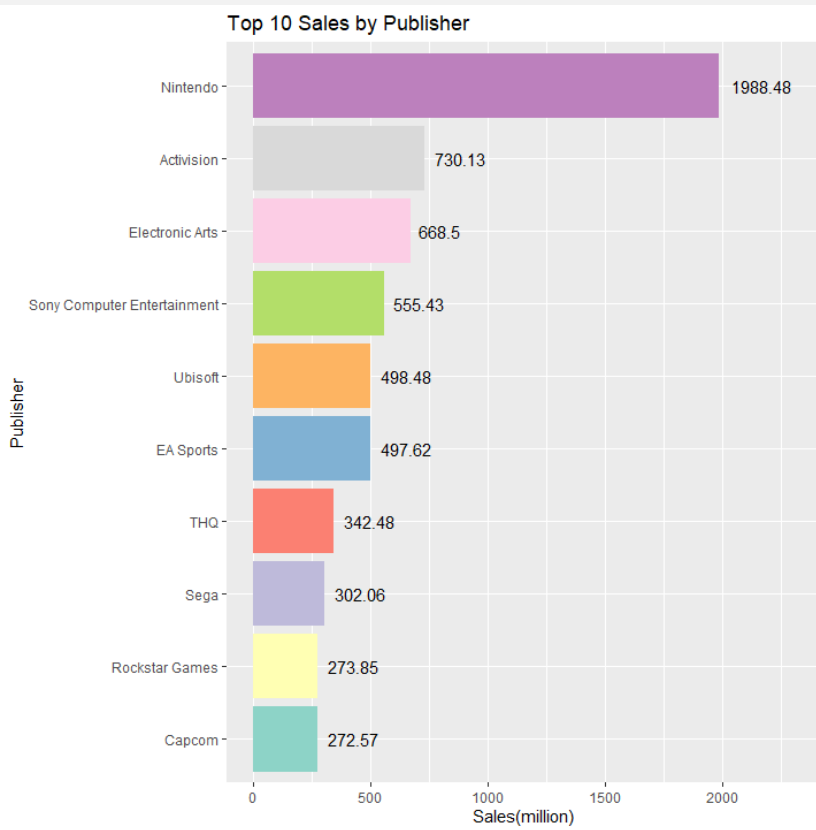
## II: New Attributes

---

```
> summarize_if(`[<-`(vgsales, vgsales==0, value= NA), is.numeric, mean, na.rm = TRUE)
  Rank Critic_Score User_Score Total_Shipped Global_Sales
1 27896.5      7.213709   8.253433      1.887258      0.3934705
  NA_Sales PAL_Sales JP_Sales other_Sales      Year salesTotal
1 0.2817352 0.1876785 0.1174917 0.06754208 2005.659 0.530876
> summarize_if(`[<-`(vgsales, vgsales==0, value= NA), is.numeric, median, na.rm = TRUE)
  Rank Critic_Score User_Score Total_Shipped Global_Sales
1 27896.5      7.5      8.5      0.59      0.14
  NA_Sales PAL_Sales JP_Sales other_Sales Year salesTotal
1 0.12      0.06      0.05      0.02 2008 0.16
```

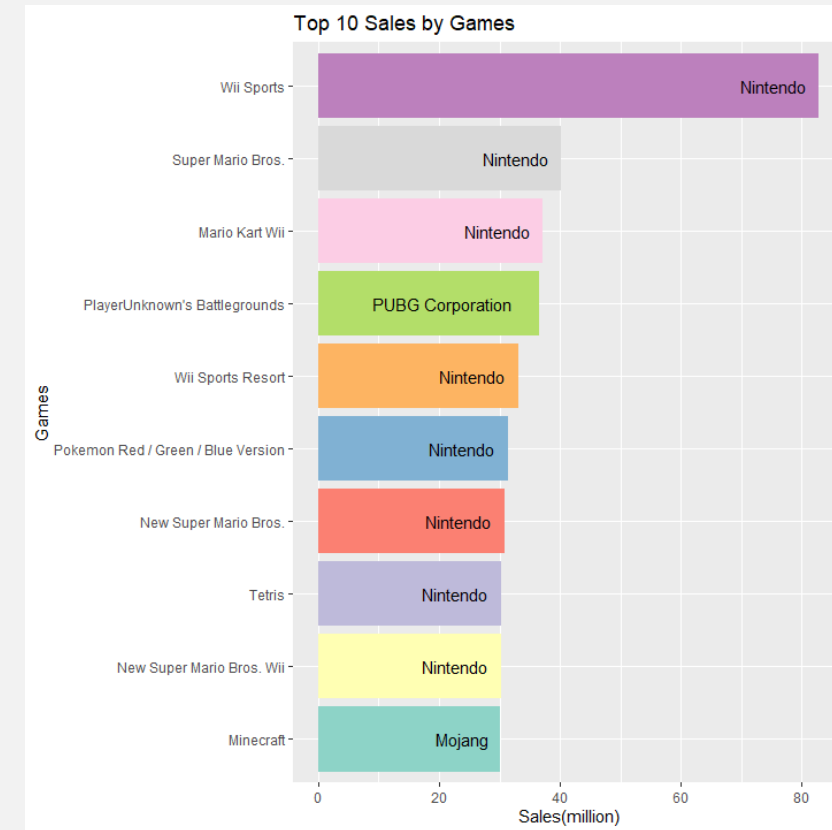
- As we mentioned previously, there were “Total\_Shipped” and “Global\_Sales” which actually got the same meaning, so we combined into a new attribute called “salesTotal”.
- The mean of the total sales data is 0.53 million, and the median is 0.16 million.

# III: Dataset Observation



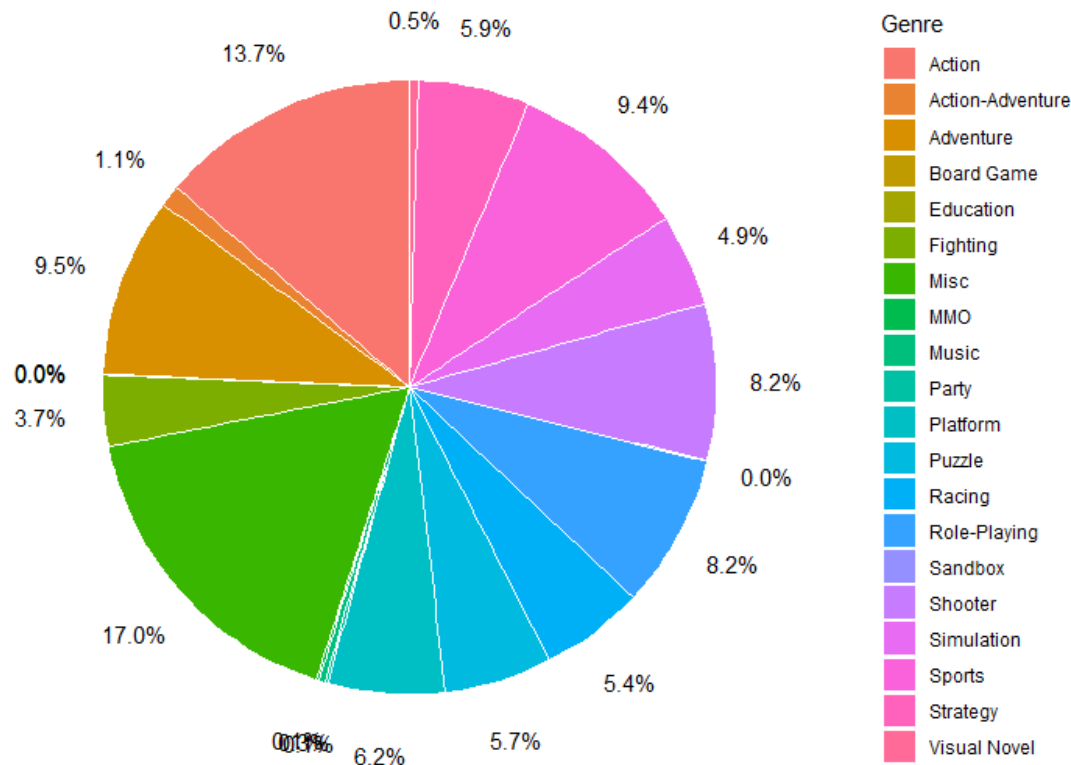
## Who is the best sales publisher?

- In this graph shows that Nintendo is obviously leading the gaming market.
- One thing that Nintendo different from the others is it contains several game studio and their games usually only sell on their own gaming console (platform).
- And if we considered the Publisher in the Top 10 Sales of Games, Nintendo covers 8 of them.



# III: Dataset Observation

Genre Percentage



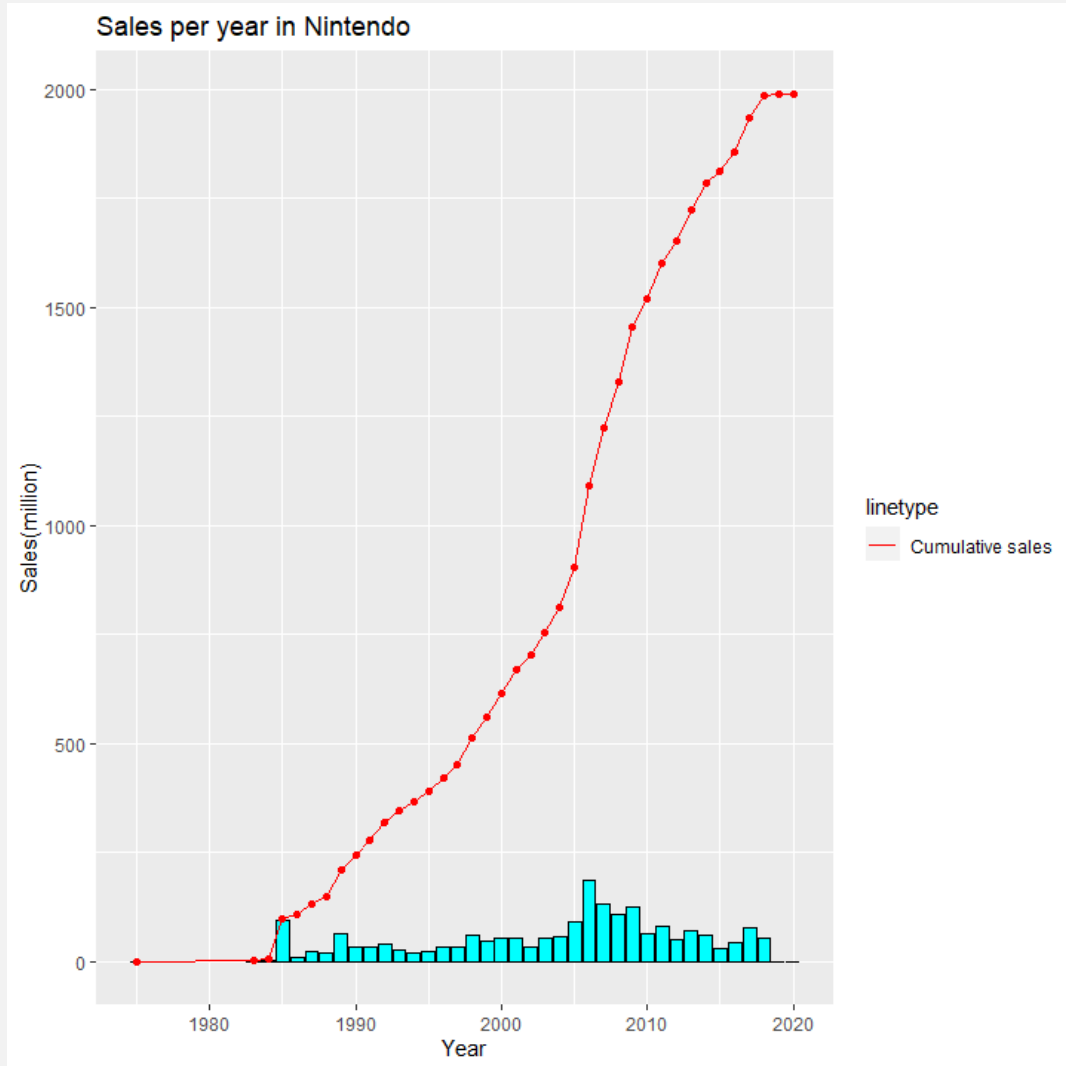
## Which genre most of the games were?

- In this graph shows that Misc and Action games is the majority of game genre in this dataset.
- Other genres excluding some unpopular kinds of games, divided quite evenly. It means that players interest is average.



# III: Dataset Observation

---



## How does Nintendo sales data like?

- In this graph shows the growth of Nintendo sales. It got a steep growth at around 2006, which is the time “Wii sport” came up.
- Apparently, we didn’t see the sales per year become more and more, that’s because the life cycle of a game usually takes a few years, caused the sales unstable.
- But if we focus on the cumulative sales, Nintendo is still growing, and overall, it’s still competitive in the games market.

# Reference

---

- *Colors (GGPLOT2). (n.d.). Retrieved October 27, 2022, from [http://www.cookbook-r.com/Graphs/Colors\\_\(ggplot2\)/](http://www.cookbook-r.com/Graphs/Colors_(ggplot2)/)*
- stefanstefan 59.8k55 gold badges2020 silver badges4242 bronze badges. (1967, September 1). *Ggplot2 geom\_text position in Pie Chart*. Stack Overflow. Retrieved October 27, 2022, from <https://stackoverflow.com/questions/62173825/ggplot2-geom-text-position-in-pie-chart>
- moodymudskippermoodymudskipper 43.7k1010 gold badges109109 silver badges154154 bronze badges. (2018, September 6). *Mean excluding zero and Na for all columns with dplyr*. Stack Overflow. Retrieved October 27, 2022, from <https://stackoverflow.com/questions/35897904/mean-excluding-zero-and-na-for-all-columns-with-dplyr>