Since the early 1980s, video games have served as an increasingly popular source of entertainment and a fixture in modern pop culture, evolving from humble beginnings to a multibillion-dollar industry today. Throughout its history, the video game market has evolved significantly, reflecting the continuous improvements to the underlying graphics and game designs as computer technology has become more powerful.  With each successive year, more advanced consoles have spawned a plethora of video game titles and genres, with thousands of titles available and new developers cropping up to try and take a share of the market. Although I have played many video games for years, I have never thought about which platforms, publishers, and genres are responsible for the bulk of the sales and quality of the games we all love to play. So here in this analysis, I would like to take a deep dive into the three categories mentioned above along with the regional sales. I'll focus on these questions to better direct the focus, and shape of the analysis.

Research questions –

1. What are the sales of video games per genre?
2. What are sales of video games per platform?
3. Visualizations of platforms’ total sales and understanding each platform’s composition.
4. What are the top 5 publishers with highest global sales?
5. What is the growth of the top publisher of games over years?
6. Generate correlation table and observe the interaction between variables.
7. Observation on Year Wise Video Game Releases
8. Understanding the global sales pattern
9. How are the changes in sales over time?

**Dataset overview -**

I analyzed data scraped from VGChartz.com, consisting of over 16,500 video game titles based either on consoles or computers and ranging from 1980-2016. The data spanned across 31 platforms, 12 genres, and over 500 publishers.

**Importing data –**

data\_vg = pandas.read\_csv(“/users/raghuveeryellapantula/Desktop/vgsales.csv”)

The fields include

* Name - The games name
* Platform - Platform of the games 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.

Lets take a quick look at the data –

data\_vg.head()

Rank Name Platform Year Genre Publisher NA\_Sales EU\_Sales JP\_Sales Other\_Sales Global\_Sales

0 1 Wii Sports Wii 2006.0 Sports Nintendo 41.49 29.02 3.77 8.46 82.74

1 2 Super Mario Bros. NES 1985.0 Platform Nintendo 29.08 3.58 6.81 0.77 40.24

2 3 Mario Kart Wii Wii 2008.0 Racing Nintendo 15.85 12.88 3.79 3.31 35.82

3 4 Wii Sports Resort Wii 2009.0 Sports Nintendo 15.75 11.01 3.28 2.96 33.00

4 5 Pokemon Red/Pokemon Blue GB 1996.0 Role-Playing Nintendo 11.27 8.89 10.22 1.00 31.37

data\_vg.describe()

Rank Year NA\_Sales EU\_Sales JP\_Sales Other\_Sales Global\_Sales

count 16598.000000 16327.000000 16598.000000 16598.000000 16598.000000 16598.000000 16598.000000

mean 8300.605254 2006.406443 0.264667 0.146652 0.077782 0.048063 0.537441

std 4791.853933 5.828981 0.816683 0.505351 0.309291 0.188588 1.555028

min 1.000000 1980.000000 0.000000 0.000000 0.000000 0.000000 0.010000

25% 4151.250000 2003.000000 0.000000 0.000000 0.000000 0.000000 0.060000

50% 8300.500000 2007.000000 0.080000 0.020000 0.000000 0.010000 0.170000

75% 12449.750000 2010.000000 0.240000 0.110000 0.040000 0.040000 0.470000

max 16600.000000 2020.000000 41.490000 29.020000 10.220000 10.570000 82.740000

**Data Cleaning:**

Counting the missing values.

data\_vg.isnull().sum()

Rank 0

Name 0

Platform 0

Year 271

Genre 0

Publisher 58

NA\_Sales 0

EU\_Sales 0

JP\_Sales 0

Other\_Sales 0

Global\_Sales 0

We have 16598 records in our data. Out of that 271 records are missing from Year and 58 records from Publisher.

We will drop these records because removing them won't disturb our EDA.

data\_vg.dropna(axis=0, inplace=False)

Now that we have eliminated the data with null values. Lets explore the year data

Lets explore thus variable a bit further.

year\_data = data\_vg['Year']

year\_data.max()

Max Year Value: 2020.0

This is an anomaly as the downloaded data is till Year 2016 only. We will remove the row(s) with wrong year, or we will try to find the real year for those columns.

data\_vg[data\_vg[“Year”] > 2017]

Rank Name Platform Year Genre Publisher NA\_Sales EU\_Sales JP\_Sales Other\_Sales Global\_Sales

5957 5959 Imagine: Makeup Artist DS 2020.0 Simulation Ubisoft 0.27 0.0 0.0 0.02 0.29

There is only one record. We can delete this.

data\_vg = data\_vg[data\_vg["Year"] < 2017]

Now that the data looks to be in good shape, we can proceed with our EDA.

Generating a correlation table to observe the interaction between variables.

Rank Year NA\_Sales EU\_Sales JP\_Sales Other\_Sales Global\_Sales

Rank 1.0 0.2 -0.4 -0.4 -0.3 -0.3 -0.4

Year 0.2 1.0 -0.1 0.0 -0.2 0.0 -0.1

NA\_Sales -0.4 -0.1 1.0 0.8 0.5 0.6 0.9

EU\_Sales -0.4 0.0 0.8 1.0 0.4 0.7 0.9

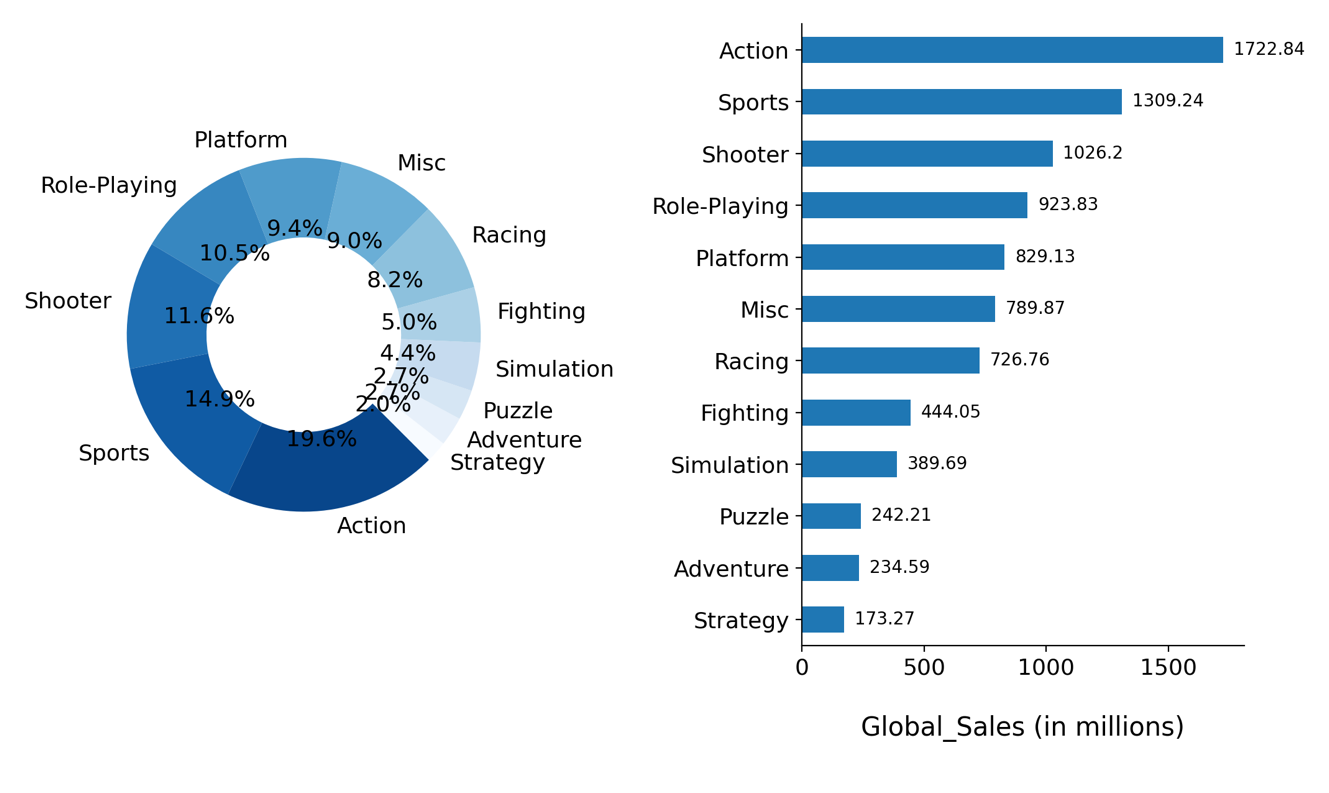
JP\_Sales -0.3 -0.2 0.5 0.4 1.0 0.3 0.6

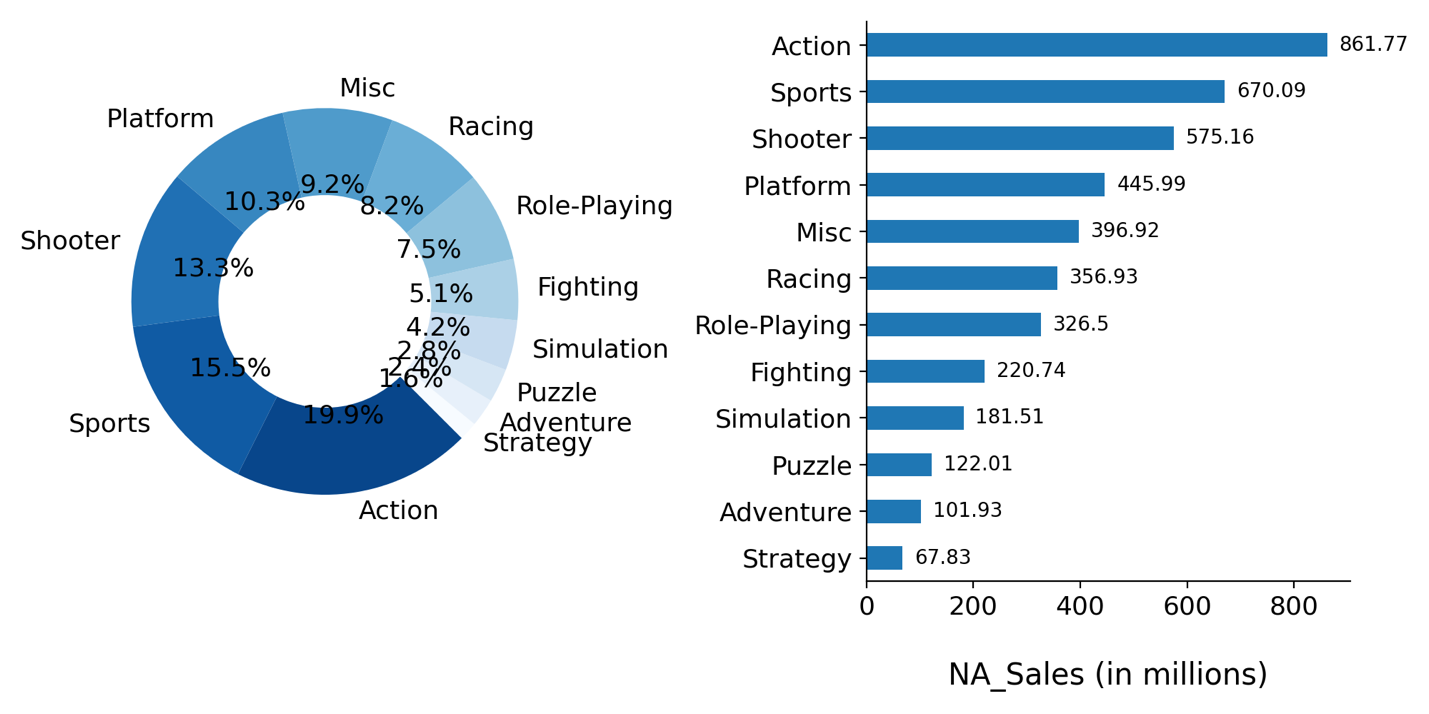
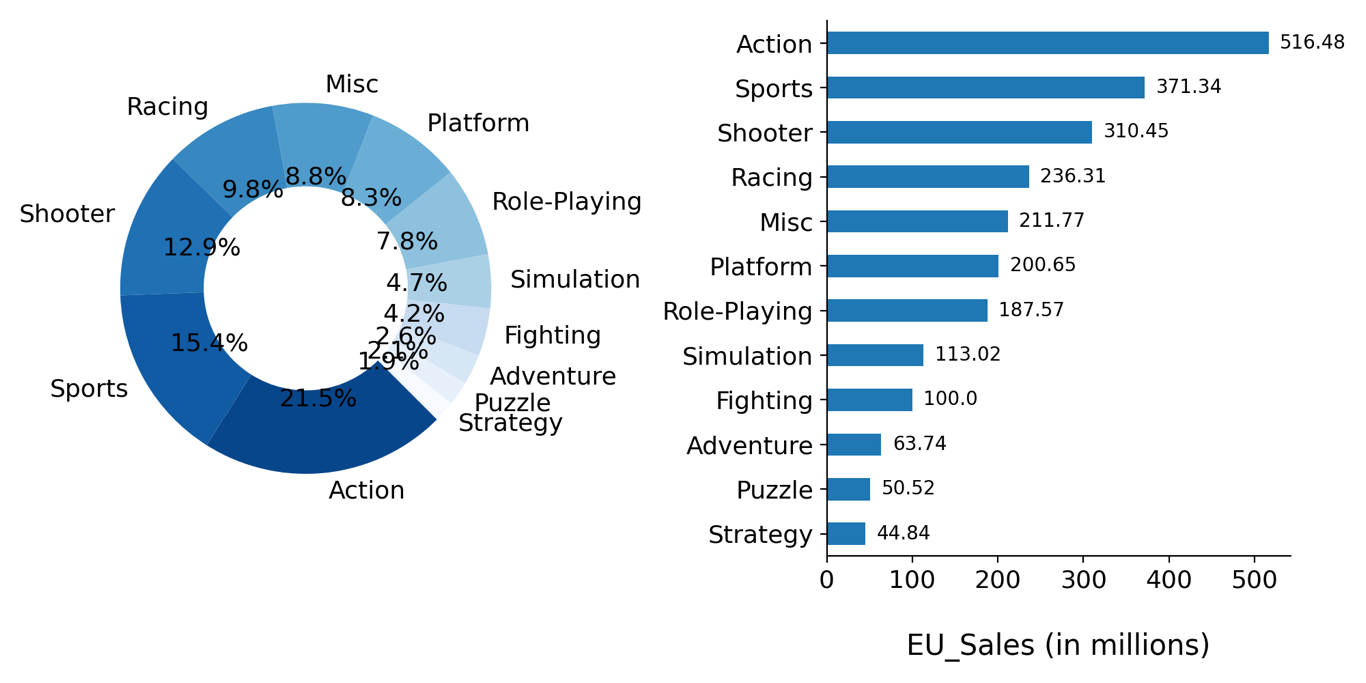
Other\_Sales -0.3 0.0 0.6 0.7 0.3 1.0 0.7

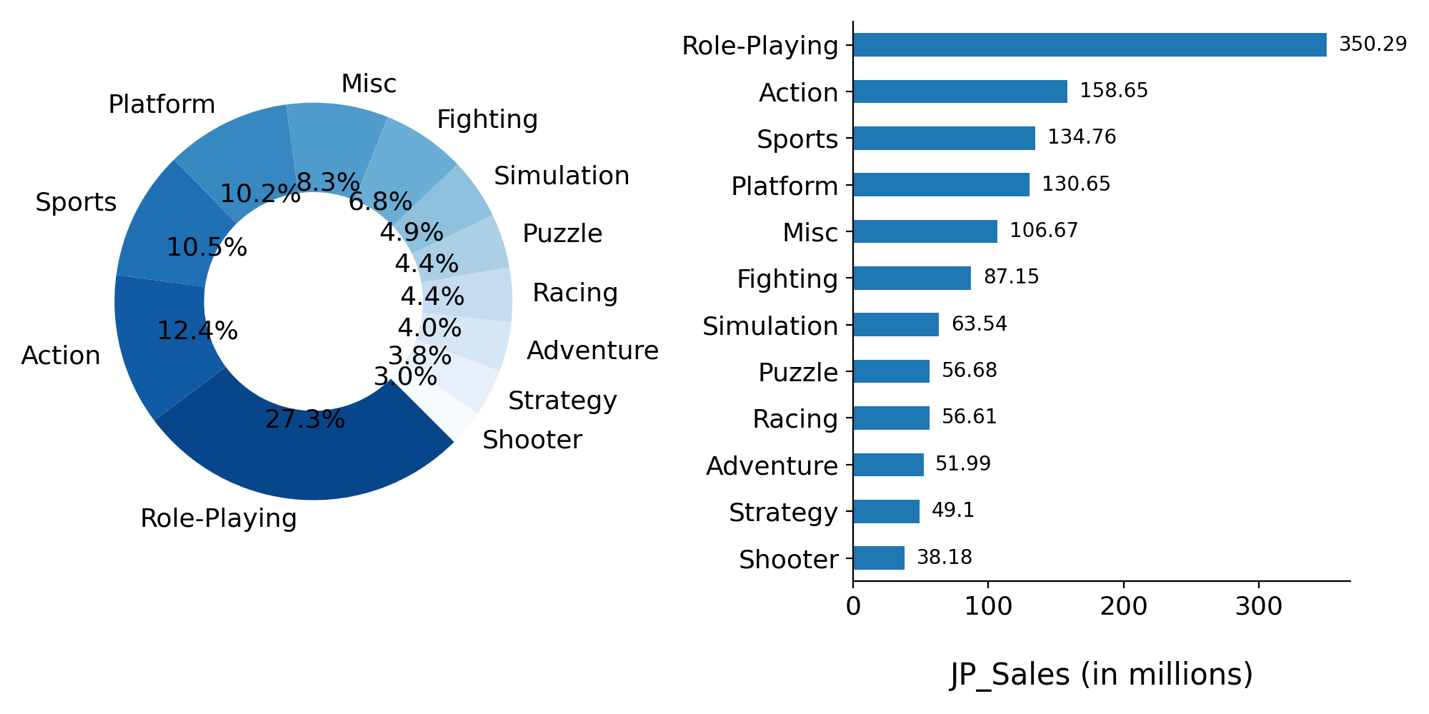
Global\_Sales -0.4 -0.1 0.9 0.9 0.6 0.7 1.0

**Sales of the video games per genre**

**Below we can see the distribution of sales of video games by genre using donut graphs and bar plots.**







**Observations -**

From the above graphs the following points can be summarized.

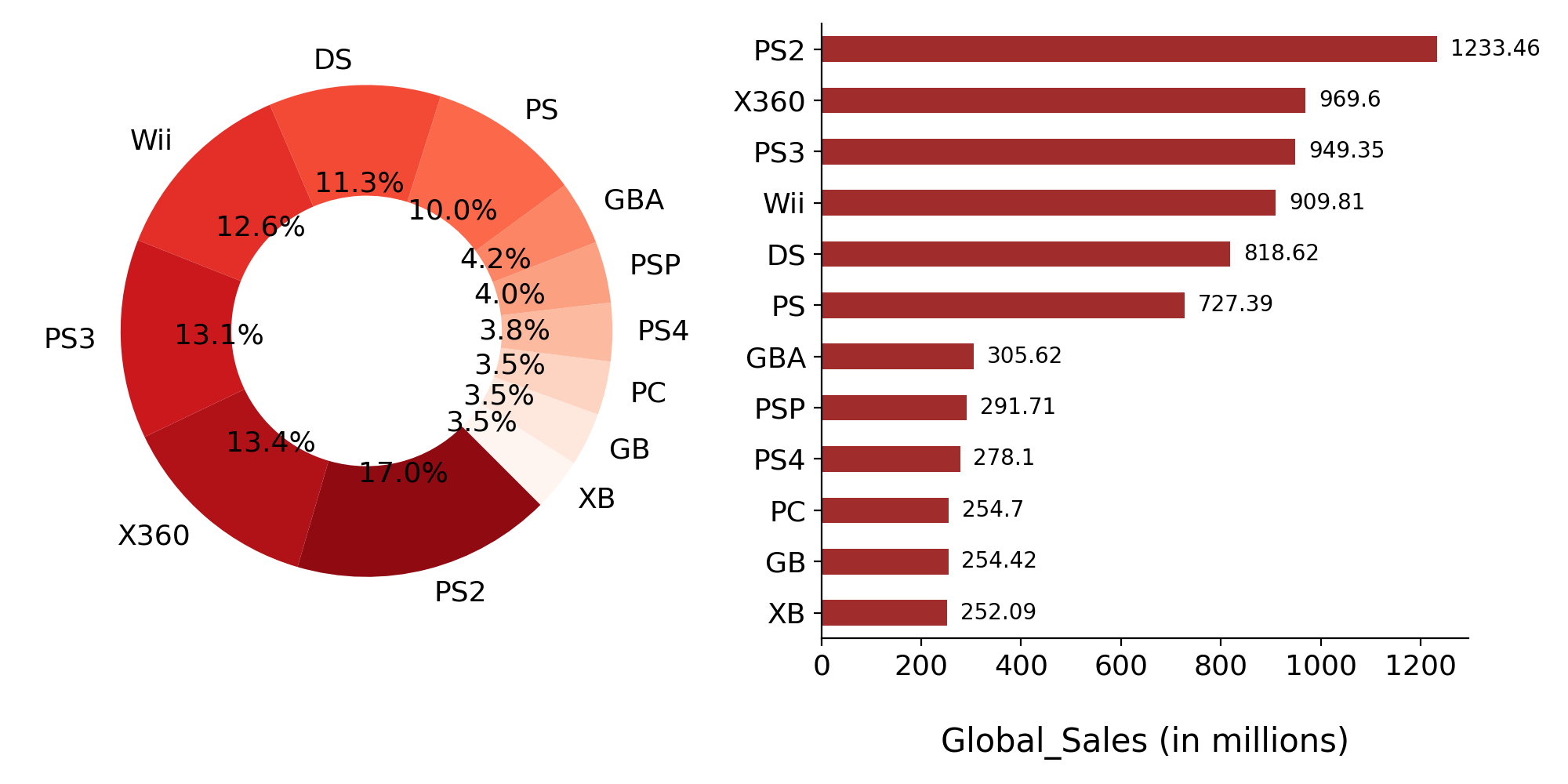
World-wide the most popular genre is Action followed by Sports and Shooter.

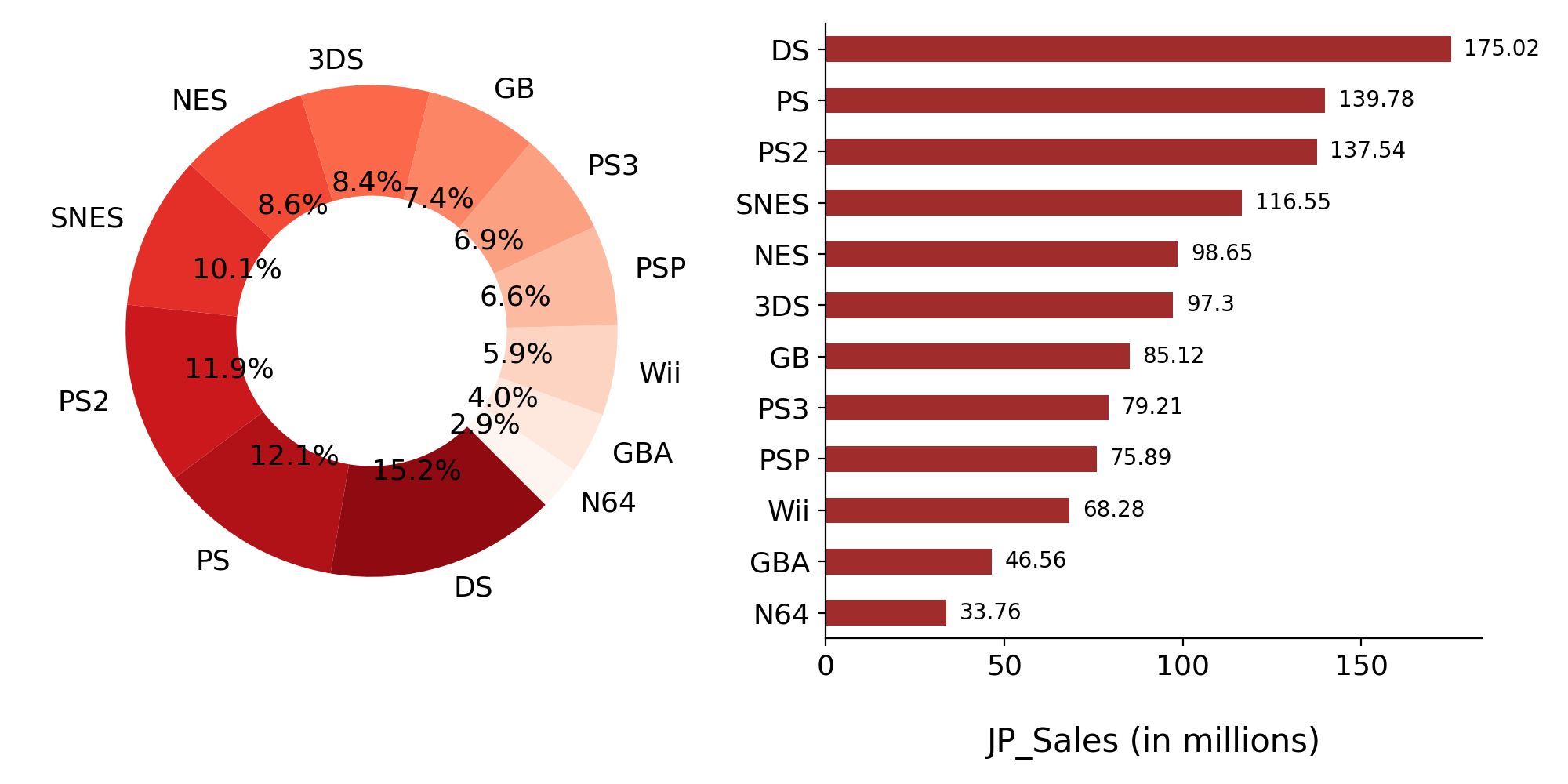
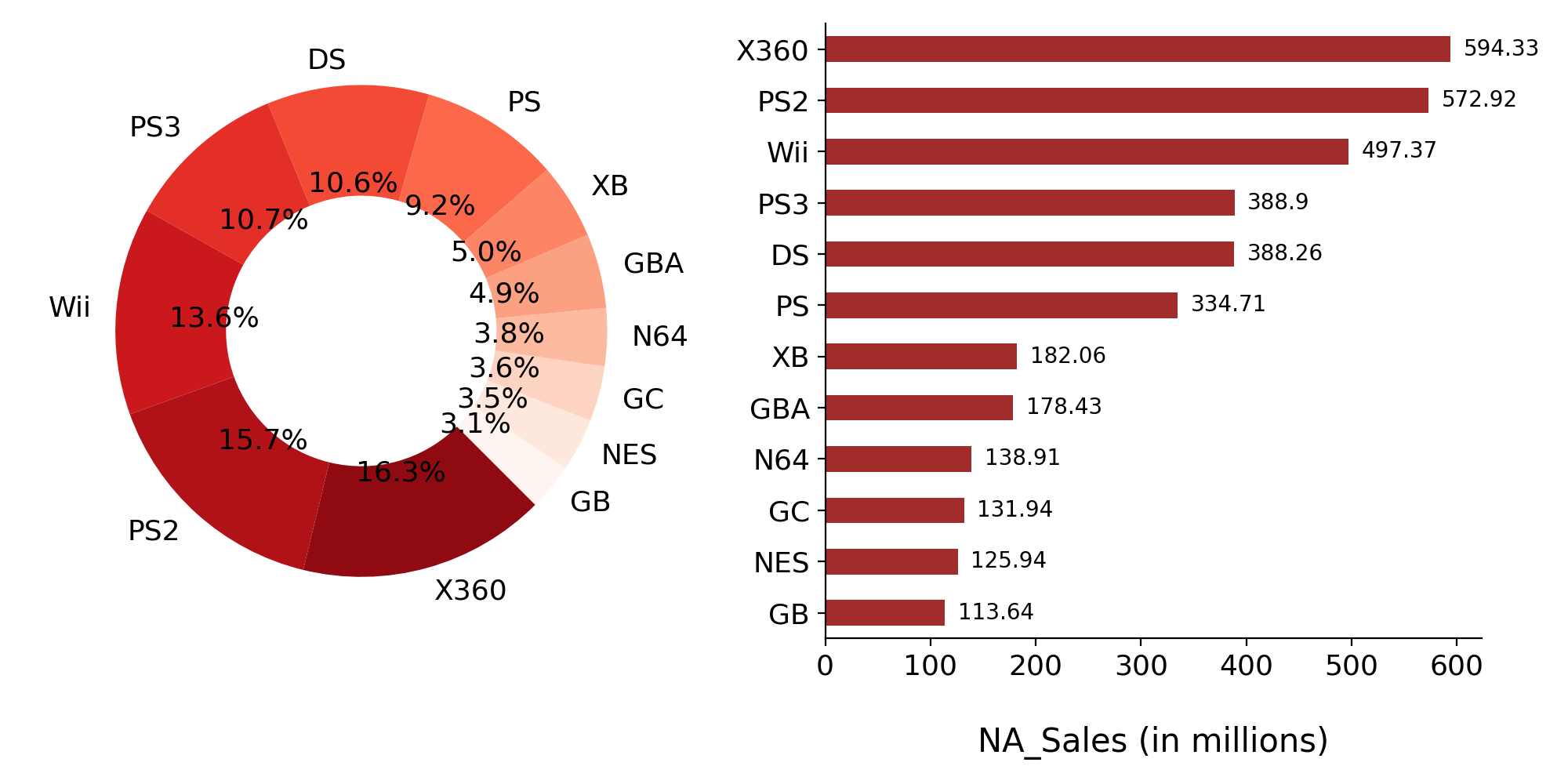
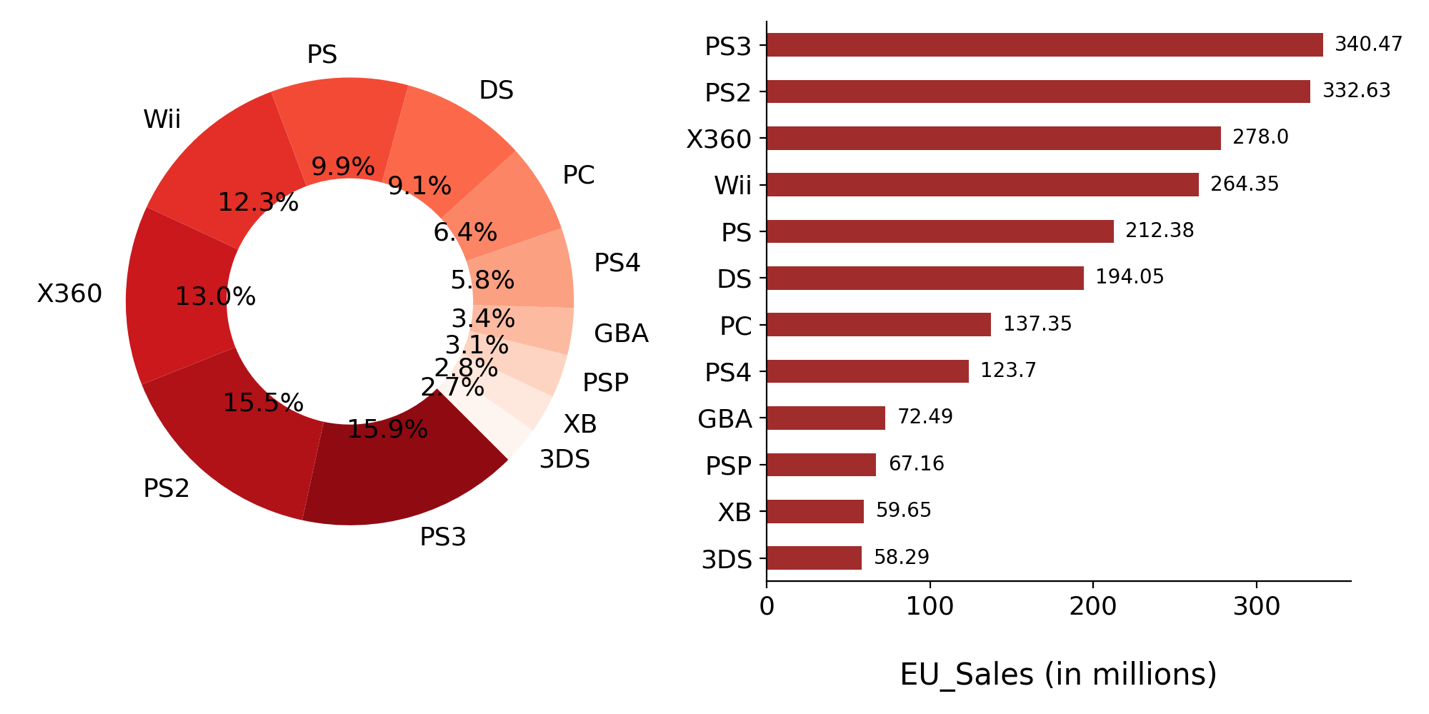
Action is the most popular genre across Europe and North America.

The most popular genre in Japan is Role-Playing.

**Sales of video games for each platform :**

Below are the platform wise distributions of video game sales globally and for each region.

****

****

**Observations**

From the above graphs the following points can be summarized.

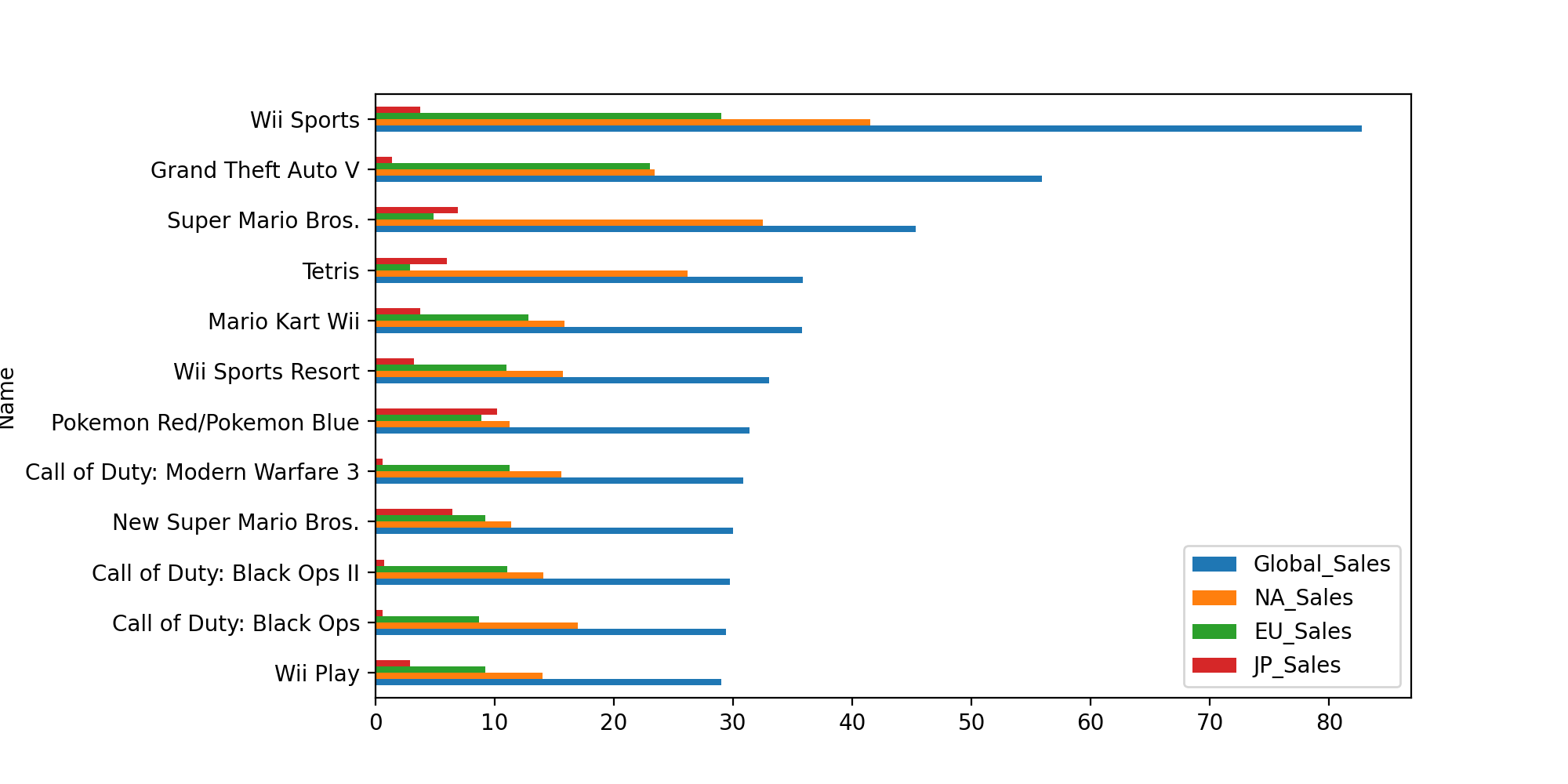
Globally the most popular platform till date is PS2(17%) followed by Xbox360 and PS3.

The most popular platform in till date North America is Xbox 360(16.3%) followed by PS2 and Wii

The most popular platform in Europe is PS3(15.9%) followed by PS2 and Xbox360

The most popular platform in Japan is DS(15.2%) followed by PS and PS2

**Sales of game titles across all regions :**

****

**Observations-**

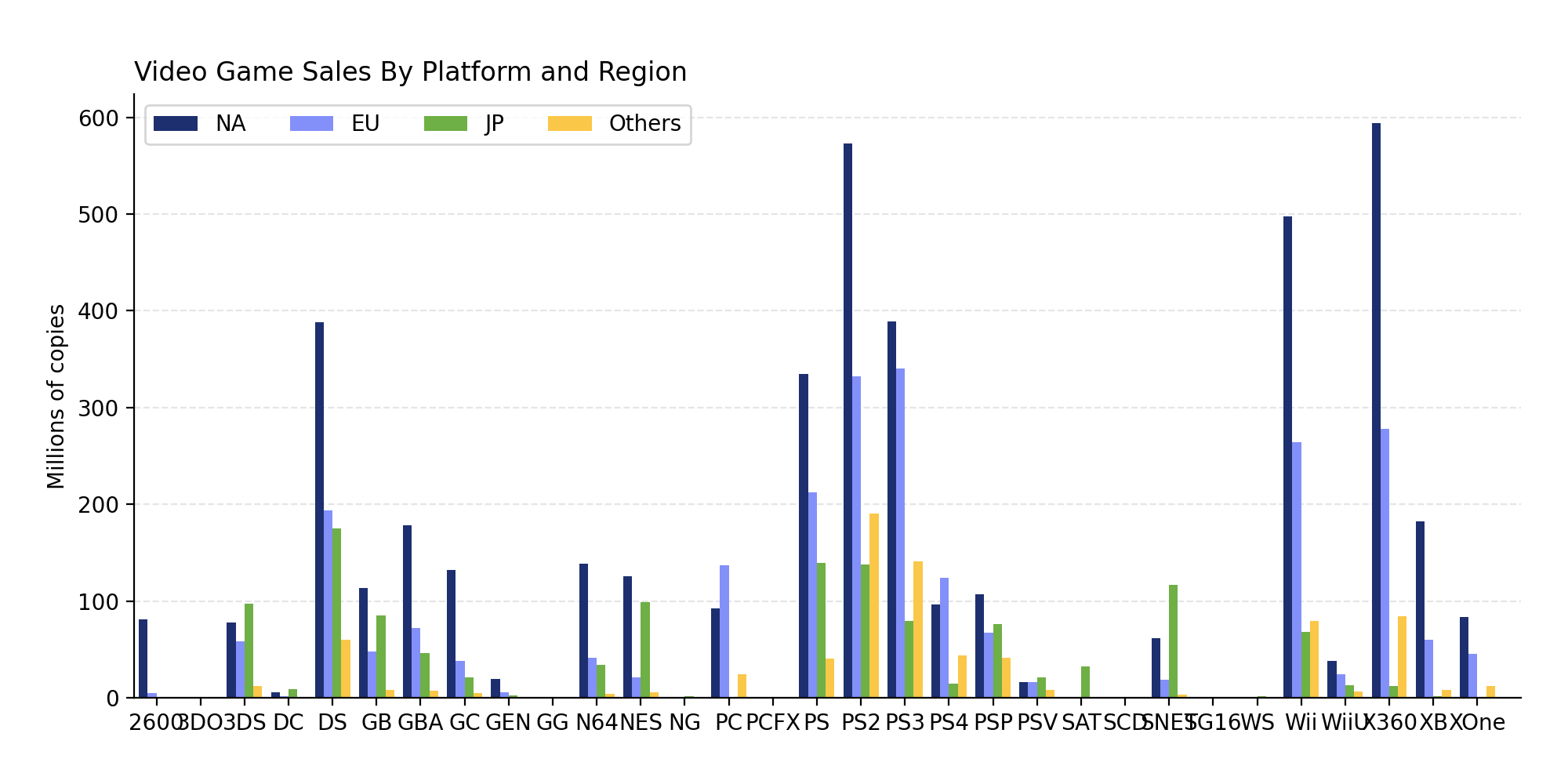
Wii Sports is the most popular in Europe, North America and also globally, where as Pokemon Red/Pokemon Blue (Role-Playing genre) is the most popular game in japan.

**Platform and region:**

I want to visualize the total number of copies sold by platform and analyze the regions where they were sold.

Having the regions already separated into columns helps a lot; we only need to group the records by

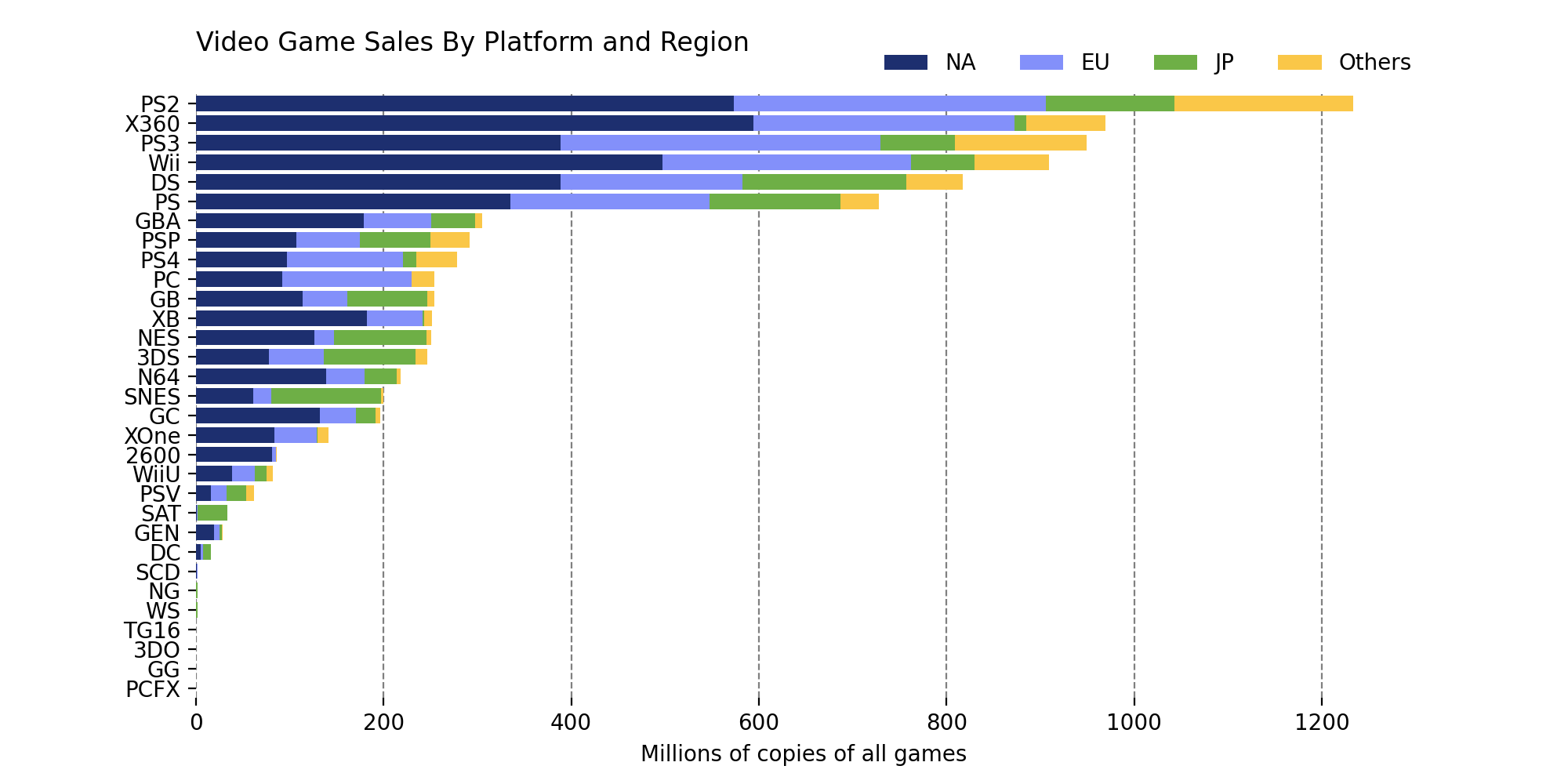
‘Platform’ and sum the values from NA\_Sales to Global\_Sales.



The above chart is hard to read. Let’s try the stacked bar chart and add a few adjustments.

First, we can sort the values before plotting, giving us a better sense of order and making it easier

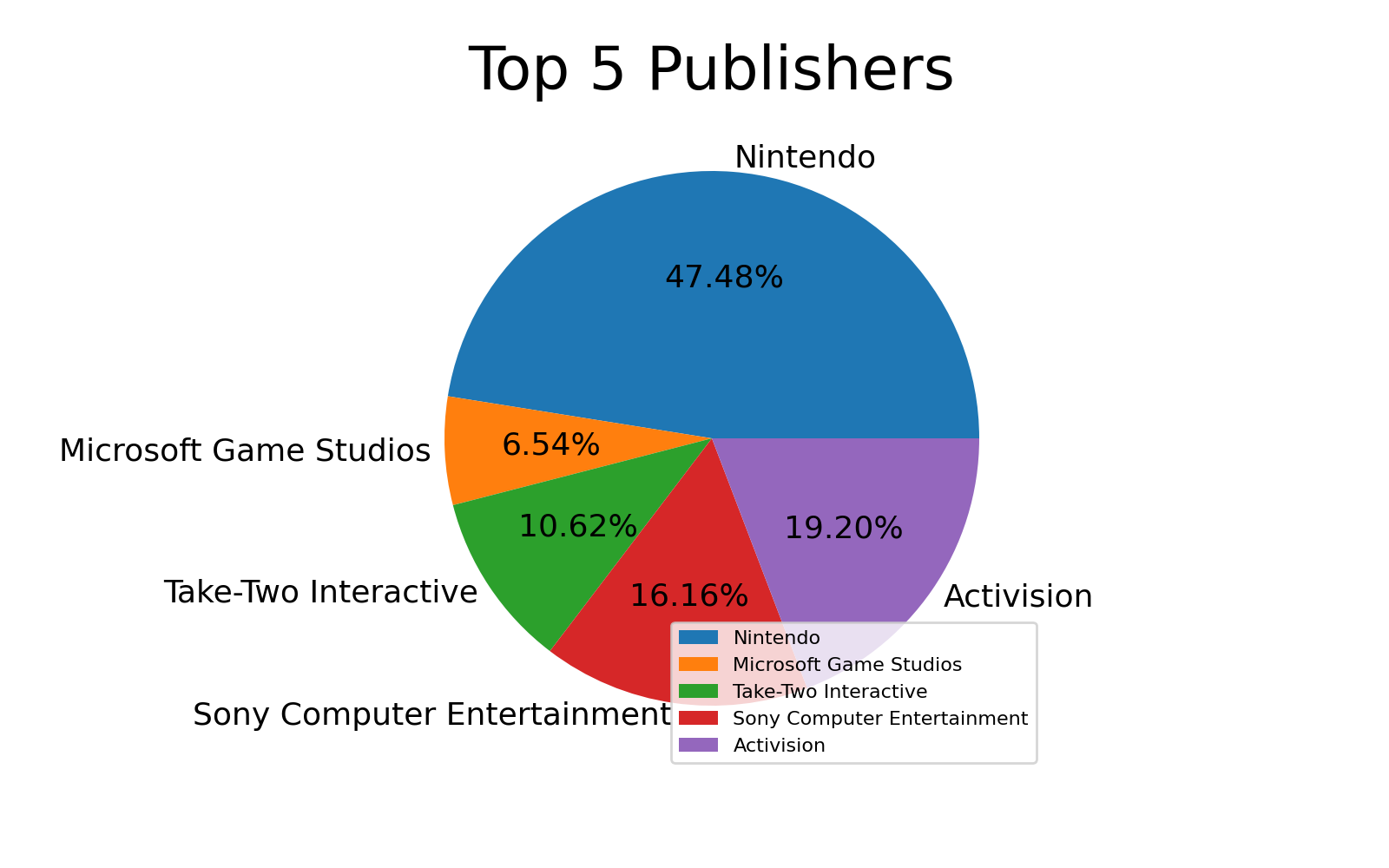
to compare the bars. We’ll do so with the ‘Global Sales’ column since it has the total.



This is way more readable than the last one. The idea here is to compare the platforms' total sales and understand each platform's composition.

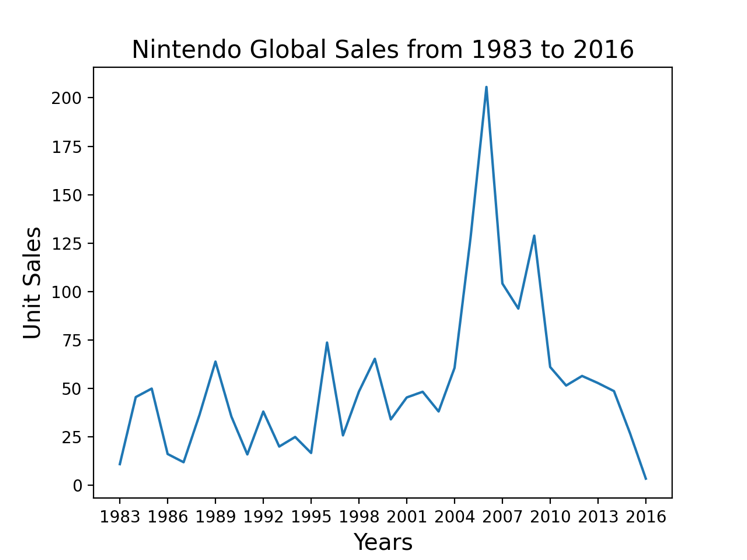
**Top 5 publishers with highest global sales:**

I want to visualize the same using a pie chart.

****

Here Nintendo takes the top spot with 47% of global sales among the top5 publishers.

**Growth of top publisher Nintendo over years ;**



Nintendo sales were at peak in the year 2006 and dropped to its lowest in 2016.

**Year Wise Video Game Release Count :**

For this I wanted to try my hands on plotly which I recently started working on.

Chart, bar chart, histogram

Description automatically generated

The plot says that the year 2009 has the highest number of games released.

**Global sales in millions:**

Teams

Description automatically generated with medium confidence

As shown in the figure, most of the sales fall within the value range of 10. We will use that information with the next plot.

**Changes in sales over time:**

**Chart, histogram

Description automatically generated**

As we can see from the plot, after a small dip in sales 2003/2004, sales exploded and peaked

2008-2010. Also known as the golden age of gaming where many games were released and sales were at its peak in the same time. The sales volume in a year is often proportional of the games released in the given year. However, one outlier being 2004, where in my opinion, the resale/collectivity of games began to gain traction.

I can say that it would be a great addition if critic scores, ESRB rating are included in the dataset. With that said, I feel that having ESRB ratings would have helped us better understanding on the sales.

Before the start of the analysis, I assumed that Microsoft was top in platform market. But to my surprise Nintendo topped the charts.

From the above analysis I would like to summarize my findings here.

When it comes to the platform market, Nintendo is a powerhouse. What surprises me the most, it's not only the quantity of games they are releasing year to year that are driving the results. Nintendo shows up more in

the top 5 sales, top 10 sales, and top 100 sales than any other platform by a long shot. The only competitor being Microsoft, due to how big of a hit the xbox 360 was.

It seems to be a recurring theme in this analysis. Nintendo has proven year after year, the quality of their games are top notch. However, you have to give credit where credit is due. Activision shows the ability to make quality games. With Call of Duty releasing every year, and a younger generation that doesn’t care as much about story lines, and single player game modes, the online realm seems to be dominated by activision and EA.

The global sales started to rise in the early 90's and for the most part, didn’t slow down(except for 2001/2002) until 2011. 2009/2010 were by far the golden era for video game. Not just for the companies behind the platform, or the studio's responsible for the product, but also for the consumer. During 2009, blockbuster games like Call of Duty Modern Warfare 2, littlebigplanet(psp),and assassins creed 2 were all very good games, and were marketed extremely well leading up to the 2009 release season.