DSCI 521 Project Phase I (Group 12)

1. A background report on the team's members, their self-identified skills, and individual contributions

Kim Dineen: Currently a Data Science major. I have Economics bachelors and an MBA with concentration in Business Analytics. I am fairly novice with coding, but I am most proficient with Python. I am comfortable using Jupiter notebook and I frequently work in tableau at work. I will use my business analytics background to help with data analysis.

Yi Yang: I have bachelor's degree in finance and cs. I am currently an MS CS student. I have one year of software development experience. I have participated in some full-stack projects which are cloud services and cluster server construction. The programming language I am most proficient in is JAVA. Although python is not the programming language, I am most proficient in, I will use my programming skill to construct an analysis model for the team to provide data results that are sufficient to support the analysis.

Dhrumil Patel: I have a bachelor's in computer science and am currently working toward my master's in data science. I have knowledge of Python, R and Java programming Languages. I am comfortable using Colab or a Jupyter Notebook as an environment. I also have some experience with visualization tools like Tableau. Additionally, I have some prior experience with data science. Additionally, through my programming skills and some prior knowledge, I can construct the data pipeline to support the project.

Arnav Goel: I have completed my undergraduate degree in computer science and done my MBA in finance. I am pursuing data science so that I can integrate data science into the finance industry.

Skills: Python, Data acquisition and preprocessing, data analysis.

Contribution: worked on part 7 and 8

1. A discussion of what you would like to your analysis to do, who/what it will support

Our analysis will use the sales data of 16,500 games to predict which game will be most profitable in the upcoming year. Using historical trends, we establish which genre and platform will top the charts as well as determine who will win this generation’s console wars. The concept of console wars has been around since the late 80s/early 90s with Sega and Nintendo battling for market dominance. Modern day is a battle between Nintendo, Sony and Microsoft and we will look at both console sales as well as single console games to determine who will come out on top. Console unique games is nothing new for Nintendo as they are a variety of series that have only been available on their consoles (example the various Mario games), however, Sony and Microsoft have only recently started implementing this strategy to get an advantage over each other. Data will show us the true impact on finances for these single console games and whether this direction is the most profitable.

1. An exhibition of analyses from dataset(s) explored, including visual analyses, captions, and useful descriptions

The dataset that was examined is represented by a few of the following exhibits:

1. What genre games have been made the most?



Fig 3.1 - Comparing games made by genre

1. Comparing Sales in different countries by genre?

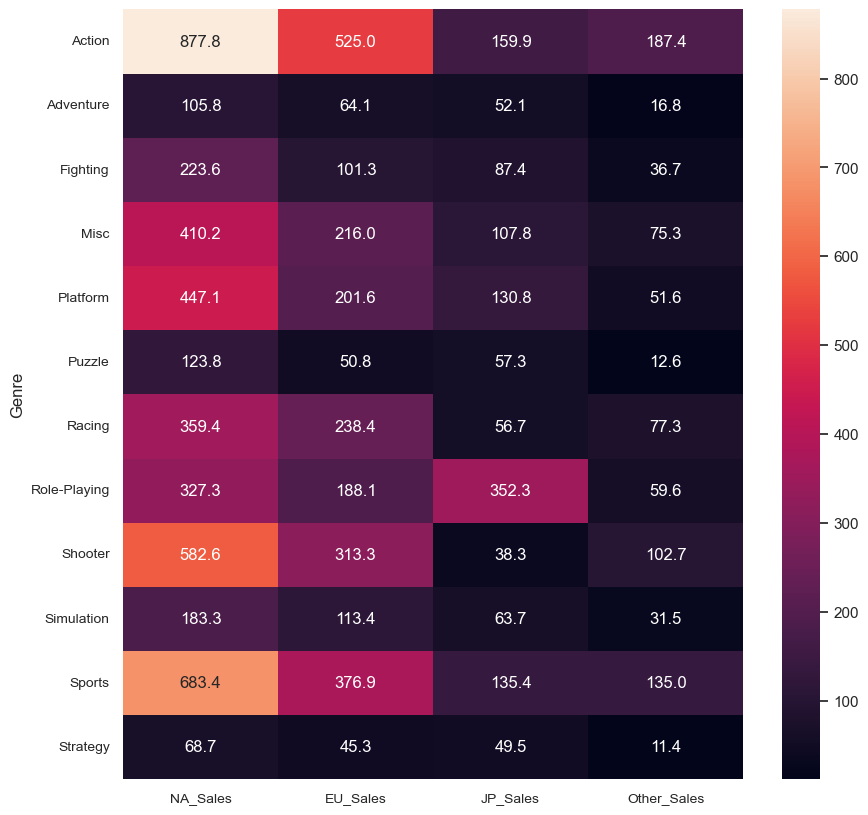


Fig 3.2 - Comparing sales by different countries by genre

1. Which genre has the highest sales price?

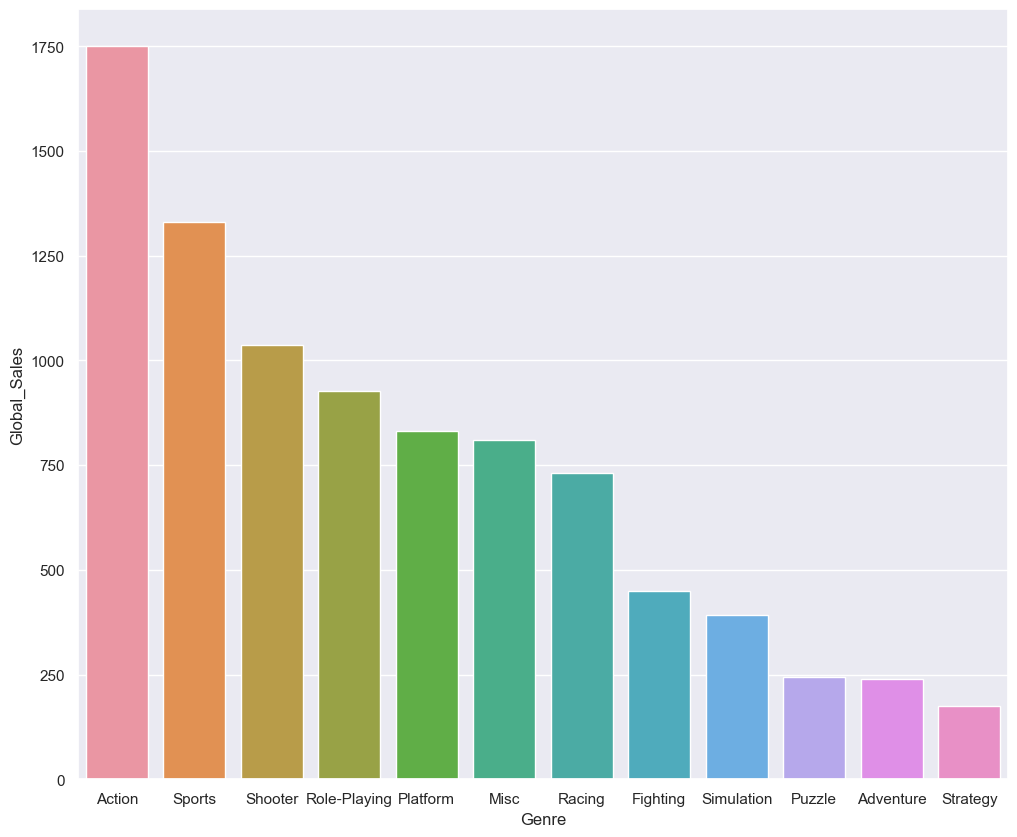


Fig 3.3 - Genre that was most popular and had highest sales price

1. Which game has the highest number of sales?

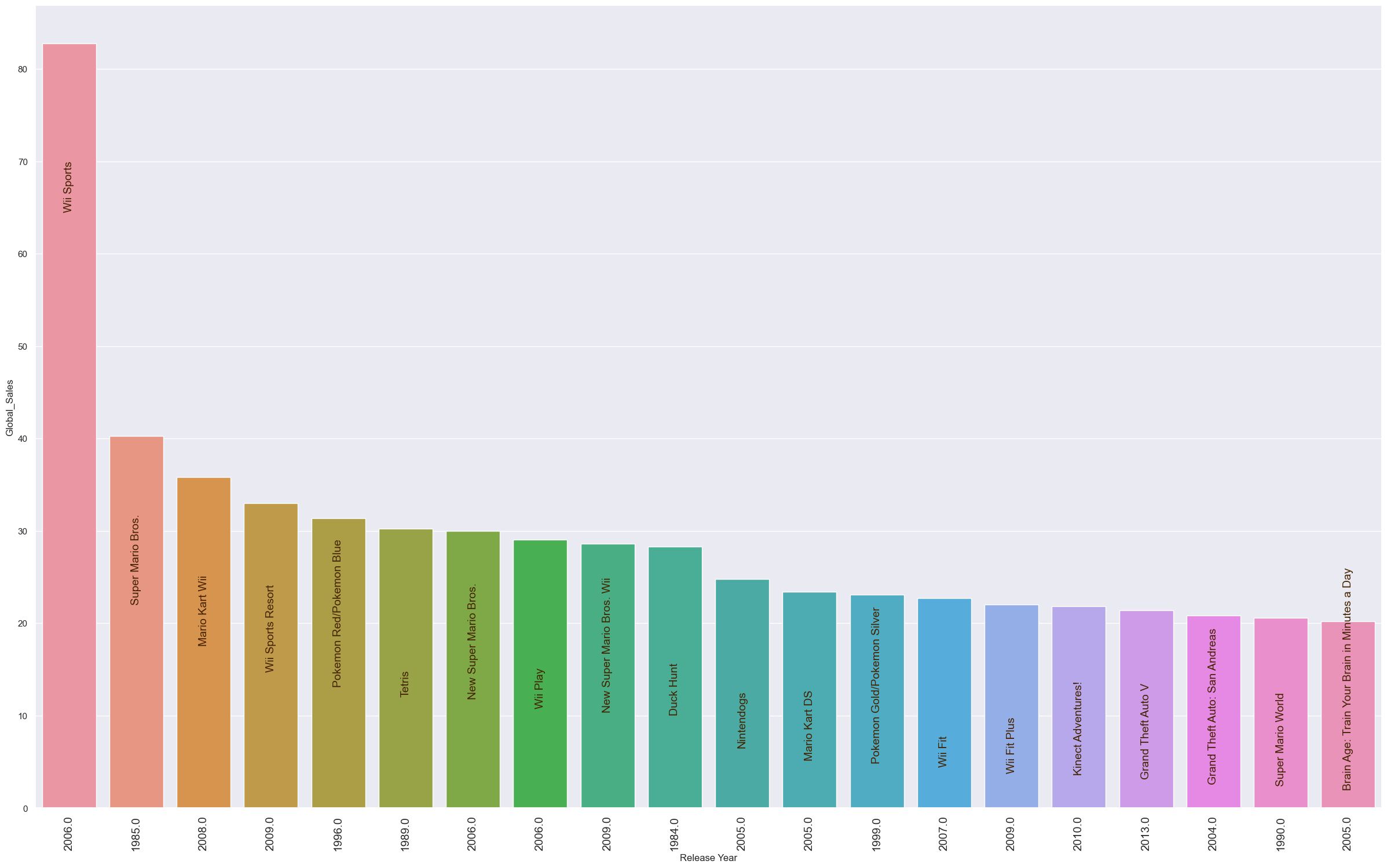


Fig 3.4 - Individual games with top sales

1. Year that saw highest sales of Games?

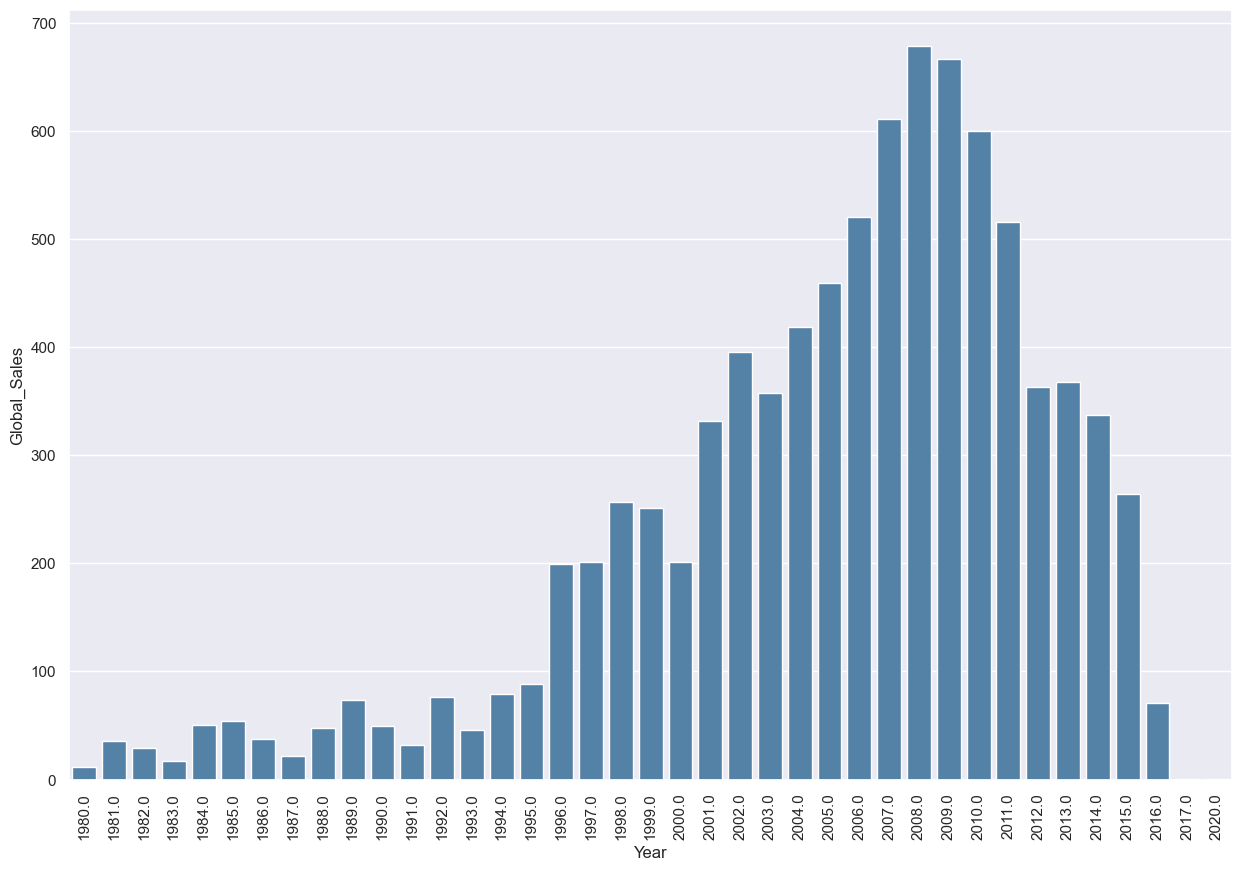


Fig 3.5 - Histogram showing the sales of games by year

1. Global Sales compared by game platform

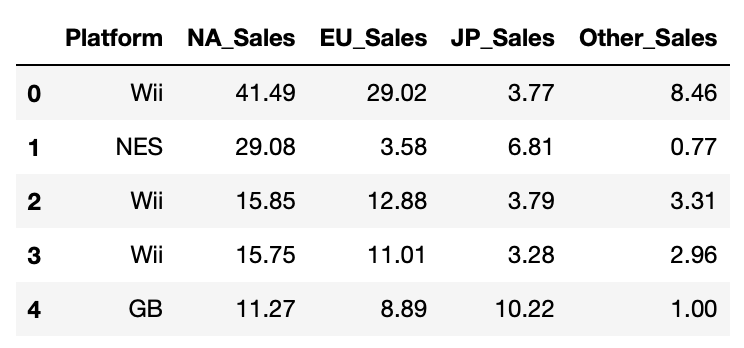


Fig 3.6 - showing global sales by different gaming platforms

1. Top Publisher every year

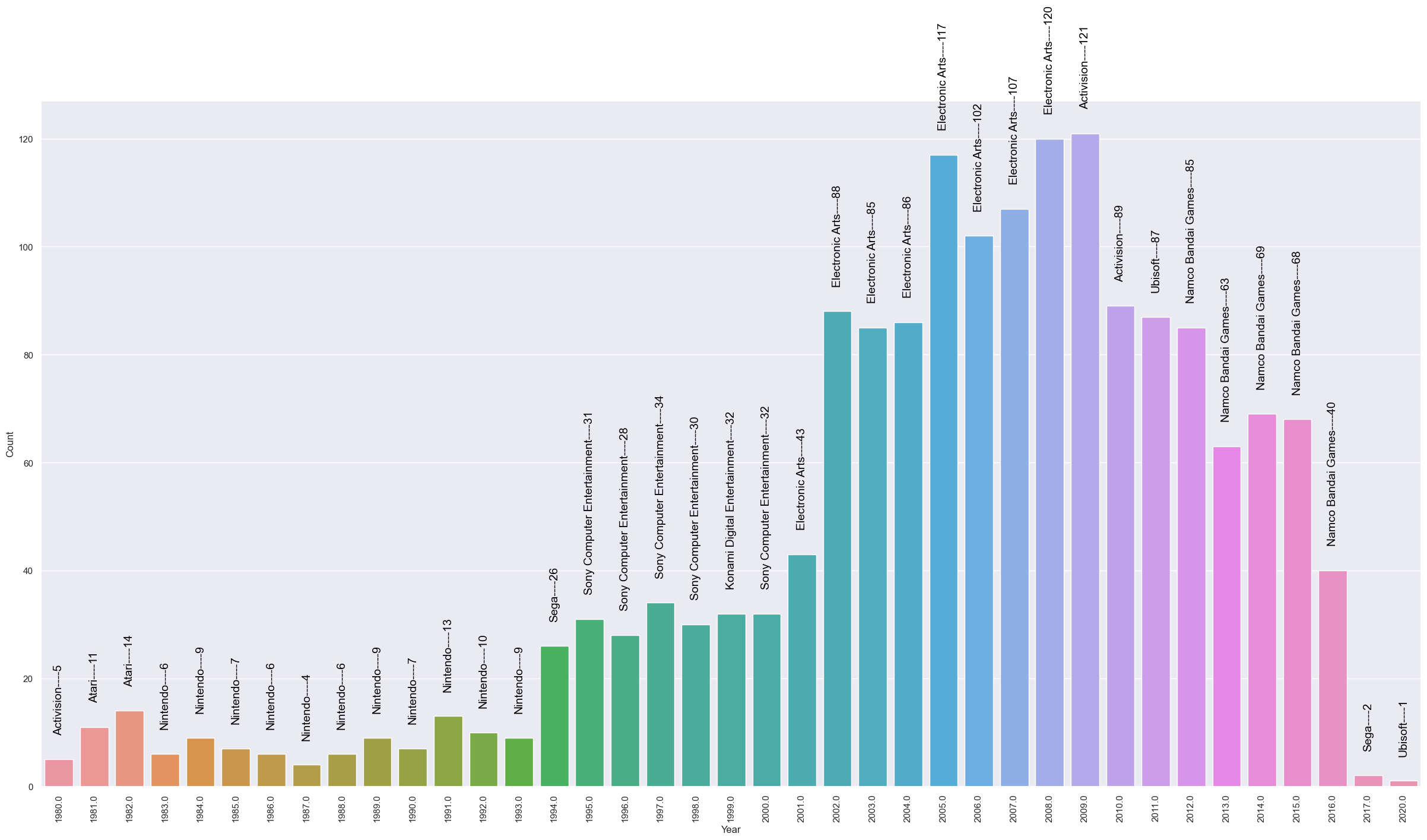


Fig 3.7 - Shows count of top publisher by years

1. A discussion of who might be interested in your analysis

Our analysis is going to cater to two general audiences. The primary audience will be gamers of all levels. Our data comes from their behaviors, so this is informing the community of their past and predicted future behaviors. This audience will be interested in the data analysis as it is talking about a topic related to one of their hobbies. The secondary audience would be stockholders of various video game companies. Our predictions are regarding the future success of upcoming videogames which will theoretically impact stock value. This data will provide a market overview at a high enough level to quickly process past and predicted video game sales.

1. A discussion of how your analysis might fit into an application or investigation

The analysis might fit into an application or investigation by comparing the game's ranking, name, genre, and release platform. The model would improve the quality of predictions for game sales. Besides, game industry investors and producers will be able to make more accurate diagnoses and predictions of video game sales. It also will assist providers in interpreting data and implementing a more predictive result.

1. A discussion of how your analysis is limited and could be improved

The limitations in the model consist of not considering all risk factors that may have a massive impact on the release of the game, and these risks may reduce the sales of the game and the purchase expectations of consumers. For example, policy and development intervals are also crucial factors affecting game releases, which will directly affect game sales and lower consumer expectations. Other limitations include not considering most of Asia, and the dataset only considers Europe, North America, and Japan. In fact, Asian regions including China, South Korea and India can also provide tremendous reference data for video game sales. In addition, the technical strength and reputation of the game company itself were not considered.

1. A selection of data for continued analysis, including justification

The selected dataset was available at <https://www.kaggle.com/datasets/gregorut/videogamesales> . This dataset contains a list of video games with sales greater than 100,000 copies. It was generated by a scrape of [vgchartz.com](http://www.vgchartz.com/).

Fields include:

1. Rank - Ranking of overall sales
2. Name - The games name
3. Platform - Platform of the games release (i.e., PC, PS4, etc.)
4. Year - Year of the game's release
5. Genre - Genre of the game
6. Publisher - Publisher of the game
7. NA\_Sales - Sales in North America (in millions)
8. EU\_Sales - Sales in Europe (in millions)
9. JP\_Sales - Sales in Japan (in millions)
10. Other\_Sales - Sales in the rest of the world (in millions)
11. Global\_Sales - Total worldwide sales.

Justification:

1. Through our analysis we are trying to answer the following questions:
2. Which region has performed the best in terms of sales?
3. The top gaming consoles are Microsoft (Xbox), Sony (PlayStation) and Nintendo, with Google acting as a new competitor. Does the dataset also back this information? Analyze with respect to different regions.
4. What are the top 10 games currently making the most sales globally?
5. What are the top games for different regions?
6. Are there any games with release year older than 2000 that are still making high sales? What are they?
7. What are the top gaming genres that are making high sales?
8. Does the publisher have any impact on the regional sales?
9. Is there any region that has out-performed global average sales?

To answer these questions, we need data referring to the sales of each region, sales by genre and many other factors which can be found in our dataset.

1. A discussion how your analysis might be completed and disseminated, i.e., who is the target audience?

After our analysis is completed, we will make our findings from our project available on GitHub, with the source code for further analysis and manipulation. The gaming companies looking to start or release a big game might find our data useful as they will know who to target, which region to target and where they should spend less. The people interested in stock investing might be interested as well as knowing the patterns in which the stocks for a certain game company rise and fall, before and after they release a big game in the market. The applications are endless with our analyzed data.