

# VIDEO GAME SALES

from 1980 to 2016

## Player 1:

Alex Nulman: 305922023

## Player 2:

Hadas Shahar: 200736858

## What?

Our dataset is a .csv file containing information about video games sales from 1980 to December 2016.

We got the data set from here: <https://www.kaggle.com/rush4ratio/video-game-sales-with-ratings>

And it can also be found on the project's Github, here: <https://github.com/shtut/VideoGames-Sales-Visualization/raw/master/Data/Video Games Sales as at 22 Dec 2016.csv>

It contains the following columns:

**Name:** The name of the game

**Platform:** The platform the game is played on (PS2, Wii, PC etc)

**Year of Release:** The year the game was released

**Genre:** The genre of the game (Action, Strategy etc)

**Publisher:** The publisher of the game (Nintendo, Sony, Ubisoft etc)

**NA Sales:** Sales for the particular game in North America, in millions

**EU Sales:** Sales for the particular game in Europe, in millions

**JP Sales:** Sales for the particular game in Japan, in millions

**Other Sales:** Sales for the particular game in other regions (not mentioned above) , in millions

**Global Sales:** Total sales for the particular game worldwide, in millions.

Each row represents a single game. The data set contains around 16k entries.

In addition, it also contained information about the **Developer**, the **Critic and User scores**, and the game's **Rating**. However, since that information was partial, we chose to ignore it.

There were also some entries with Year of Release values of 2017 and 2020 which we removed, since our data claims to contain sales only until 2016.

# Why?

Our dataset provided a lot of information and interesting ways to slice it.

While we couldn't find 1 graph that can visualize our entire dataset, we figured several targeted graphs may provide a more complete picture.

We wanted to answer the following questions:

1. What are the most popular games of all times?
  - {Lookup Extremes}
2. What are the most popular genres of all times?
  - {Lookup Extremes}
3. How has video game sales changed throughout the years? How have they changed per region?
  - {Discover Trends}
4. Have the popular genres changed over the years?
  - {Discover Trends}
5. Is there any correlation between Genre and Platform? Has this changed over time?
  - {Discover Correlation}
  - {Discover Trends}
  - {Summarize features}
6. How have the different publishers changed in popularity over the years?
  - {Lookup/Browse features}

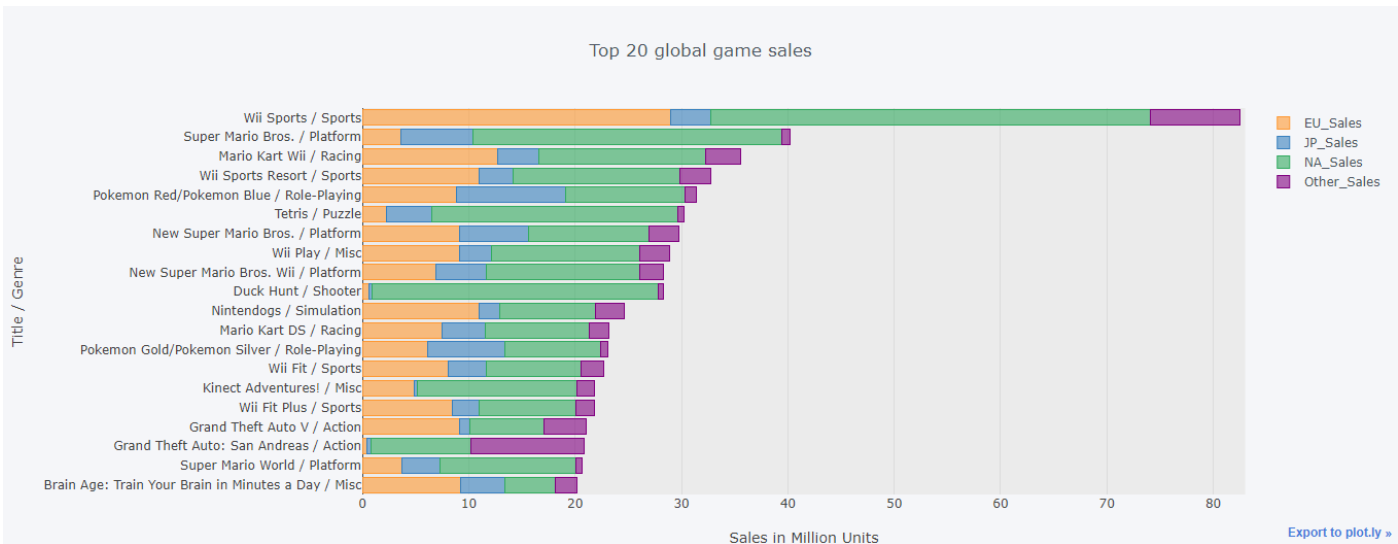
# How?

To answer the aforementioned questions, we've devised the following graphs:

1. What are the most popular games of all times?
  - **Interactive Stacked Bar chart of the top 20 games by global sales**
2. What are the most popular genres of all times?
  - **Interactive Stacked Bar Chart of all the genres by global sales**
3. How has video game sales changed throughout the years? How have they changed per region?
  - **Line Chart – each line representing a region, over time**
  - **Normalized Stacked Bar Chart of the region sales over time**
4. Have the popular genres changed over the years?
  - **Small Multiples of Area Charts per genre**
5. Is there any correlation between Genre and Platform? Has this changed over time?
  - **Heat Map with Genre/Platform by global sales**
  - **Interactive Heat Map with Genre/Platform with time slider**
6. How have the different publishers changed in popularity over the years?
  - **Heat Map of the top 20 publishers over time by global sales**

## 1. What are the most popular games of all times?

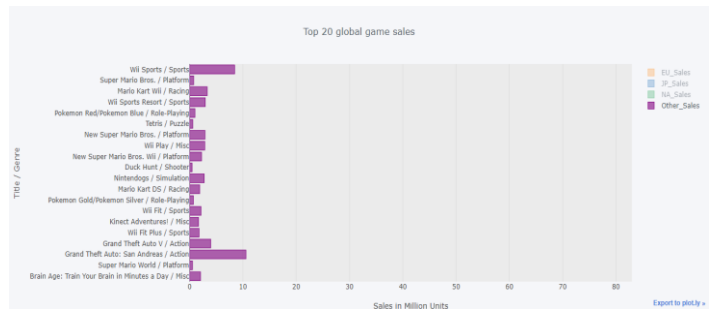
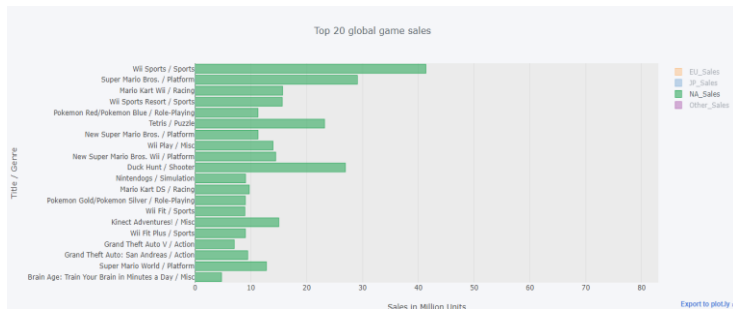
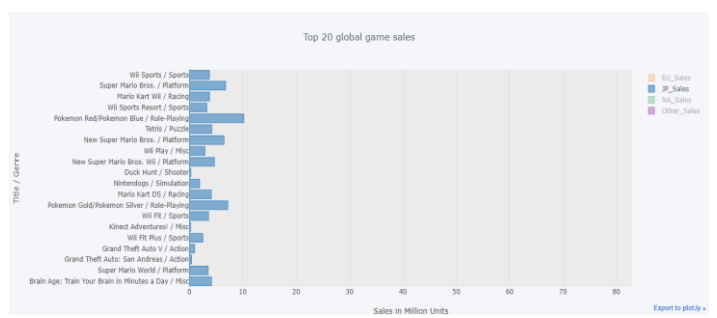
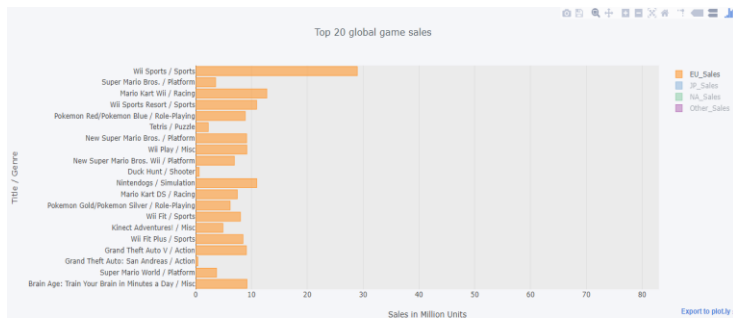
For this, we created a stacked bar chart based on global sales values, for the top 20 games:



This showed that Wii Sports stood at the top with around 80 million sales- over twice as much as 2<sup>nd</sup> place- Super Mario Bros.

We thought it was odd that Wii Sports did as well as it did considering the other names on the list were more familiar, but later we found that Wii Sports sold as part of the Wii platform, and so all its sales were just Wii sales, and not on its own merit.

With the interactive nature of the graph, it was also possible to see how the top global games fared per region:



While in Europe and North America Wii Sports wins again by a large margin- the same can't be said about Japan and Other regions.

## 1. Evaluation:

Pros:

- The graph provided a list of the top 20 games, showing the how the most popular games fared per region.

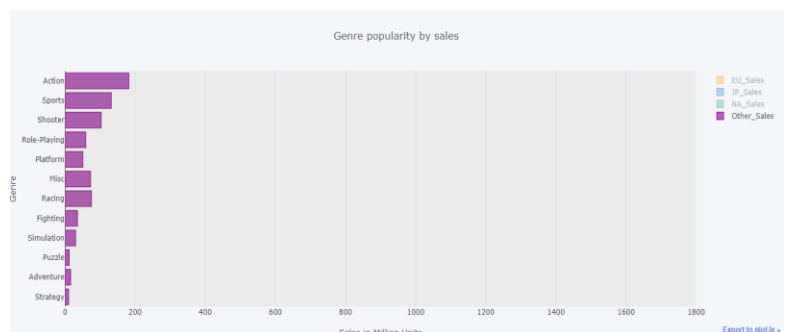
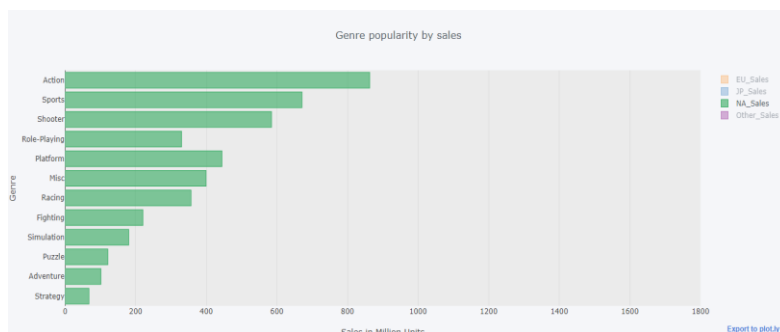
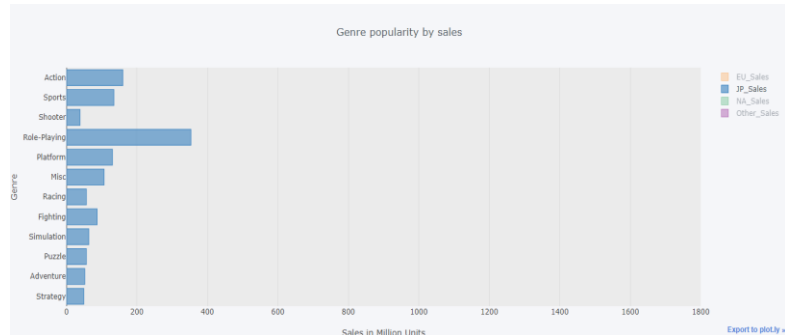
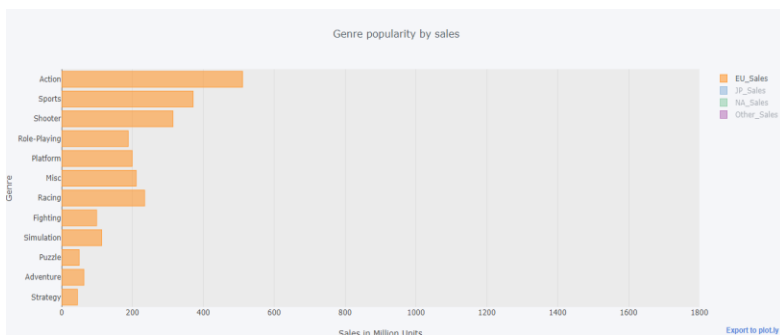
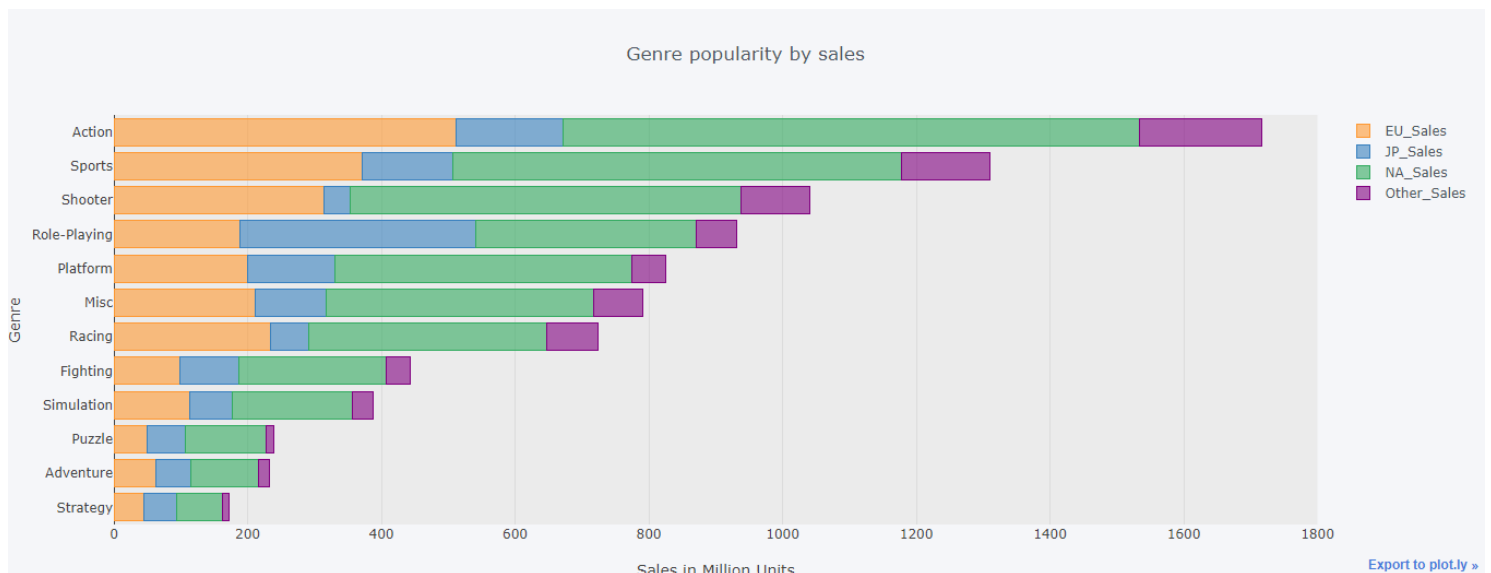
Cons:

- The graph doesn't show all games, and it's not possible to see how an individual game sold.
- Stacked bar charts are sometimes hard to instantly understand.

$$V = T(25) + I(15) + E(15) + C(15) = 70\%$$

## 2. What are the most popular genres of all times?

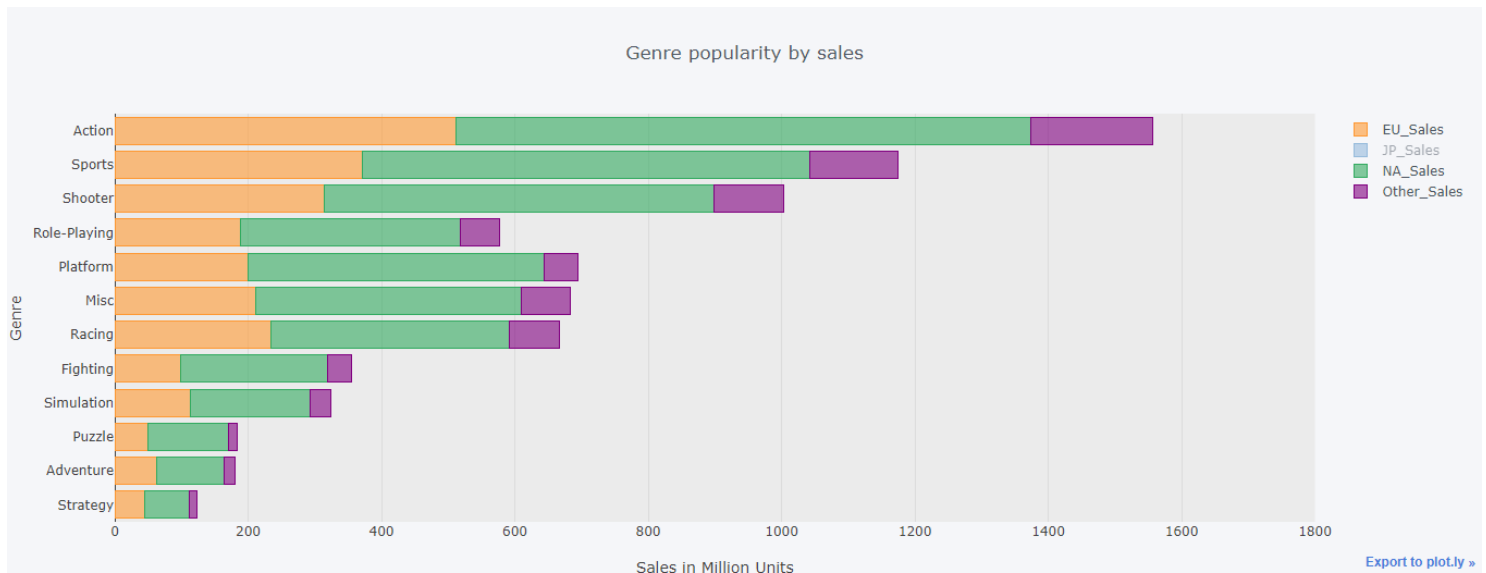
Similar to the top games graph, we figured the best way to show the most popular genres is also through an interactive stacked bar chart:



The top 3 genres – Action, Sports and Shooters, seem to be in the same hierarchy in most regions, Japan being the exception.

While Role-Playing seems to be 4<sup>th</sup> on the global scale, in Japan it is by far the most popular.

This is interesting- Japan's choice of genre influences on the top most popular genres. Without Japan's sales, our graph looks quite different:



Without Japan, Role-Playing would've been 7<sup>th</sup> on the list.

## 2.a Evaluation:

### Pros:

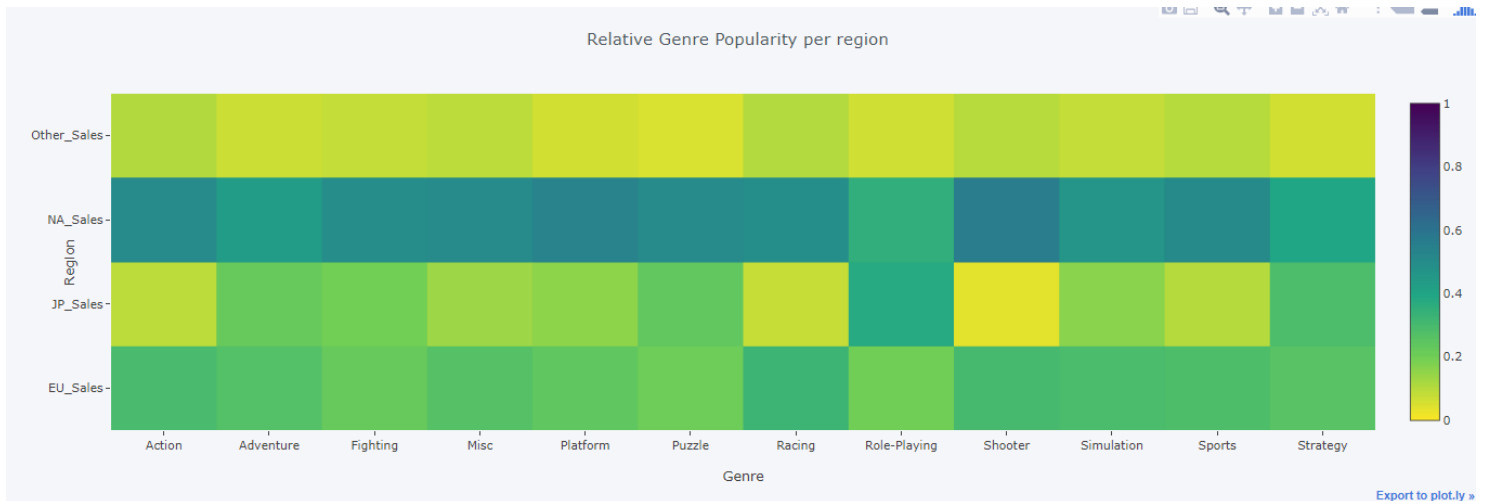
- The graph provided a list of the top genres, showing the how the most popular genres fared per region.
- Contains information on all the genres

### Cons:

- Stacked bar charts are sometimes hard to instantly understand.

$$V = T(25) + I(15) + E(15) + C(20) = 75\%$$

Another approach to this question could be done by using a heat map, where the y-axis is the region and the x-axis is the genre, normalized by the total number of sales:



This shows the distribution of each genre per region, as well as the most popular genres per region.

It also shows Japan's preference of Role Playing over North America.

## 2.b Evaluation:

Pros:

- Easier to compare each genre between regions

Cons:

- Heat map is not sorted in any way, can't see trends
- similar values create similar colors, making it harder to differentiate and evaluate which is higher

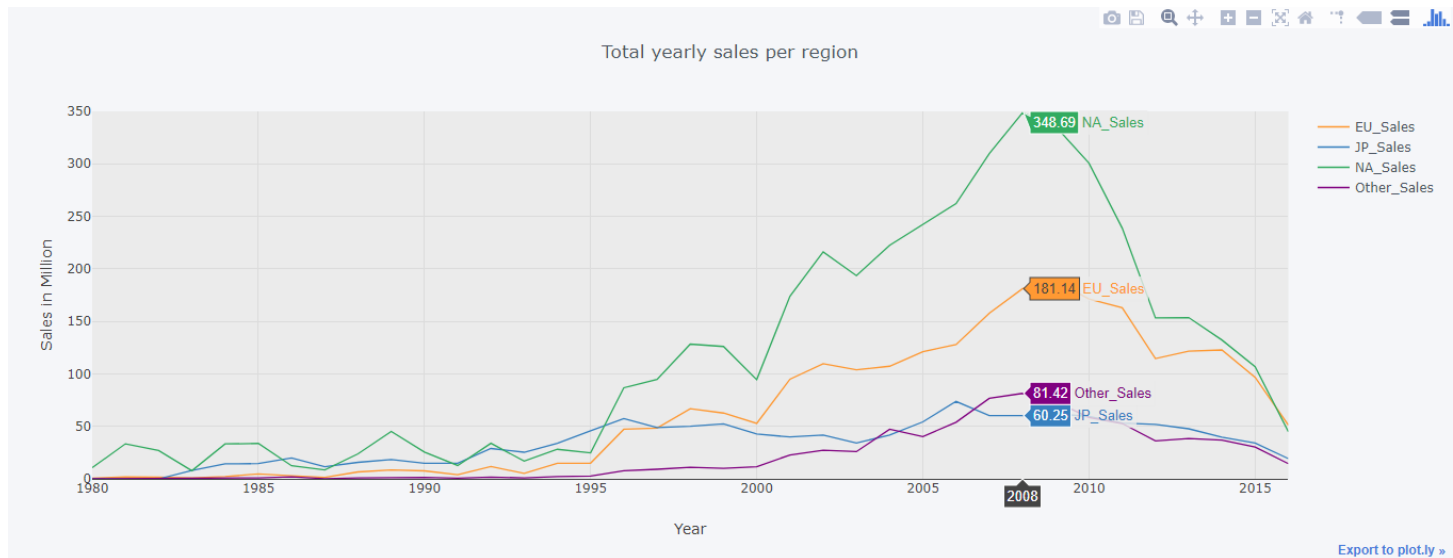
$$V = T(15) + I(10) + E(15) + C(15) = 55\%$$

### 3. How did video game sales changed throughout the years? How has it changed per region?

#### a. How did video game sales changed throughout the years?

While these graphs displayed the most popular games/genres/platforms they did not take into account the time axis in any way, so we were unable to draw any conclusions regarding trends over time.

So- putting the time (year of release) on the x-axis, we wanted to see how the sales were divided between the regions, and if that has changed in any way over time:



Here we can see that North America had the most growth over time, but also that there's a decline in sales after 2008. We tried to think why that might be, and our assumption is that it's caused by the fact that most of the games' sales occur after their year of release.

That means that some of the sales for games which were released after 2008 don't appear here, since our data only tracks the sales until 2016.

Another theory we had was the possible entry of mobile games to the market. Mobile games don't appear in our dataset but they do take up some of the sales, especially after smart phones became popular (around 2010).

#### 3.a Evaluation:

Pros:

- Simple graph which displays the total sales in every region over time- simple to compare between regions in a given year, and not too cluttered.

Cons:

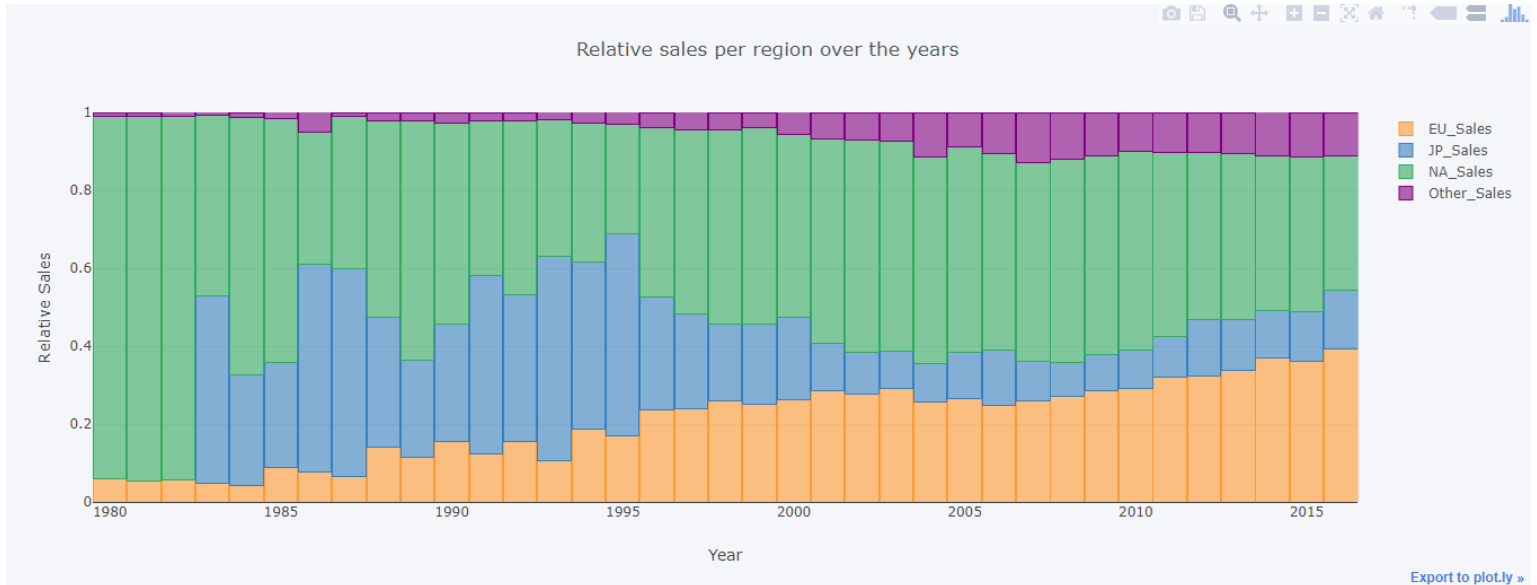
- Not very visually engaging

$$V = T(25) + I(25) + E(25) + C(20) = 95\%$$



## b. How did video game sales change per region?

While the previous graph provides good input regarding the regional sales over time, it doesn't show the relative portion of each region. For this we've created a stacked bar chart, where each column represents 100% of the sales that year:



This painted an interesting picture- while japan proclaimed the majority of sales in the 90's, their share shrunk as the western world came to see video games as legitimate pastime and not as a niche hobby.

Both Europe and "Other" have increased in sales over the years- both in market share and total sales number. North America however, while rising in absolute value, is slowly taking less of the grand total.

### 3.b Evaluation:

Pros:

- Easy to compare between the regions share per year
- Shows a trend

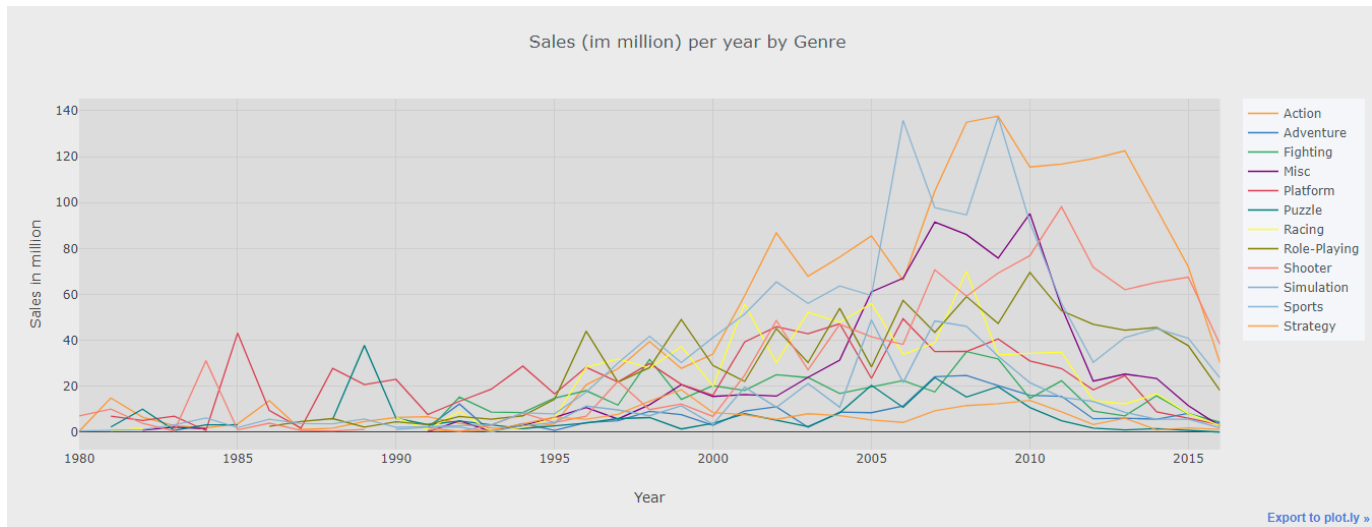
Cons:

- Stacked bar charts are sometimes hard to instantly understand.

$$V = T(20) + I(25) + E(25) + C(20) = 90\%$$

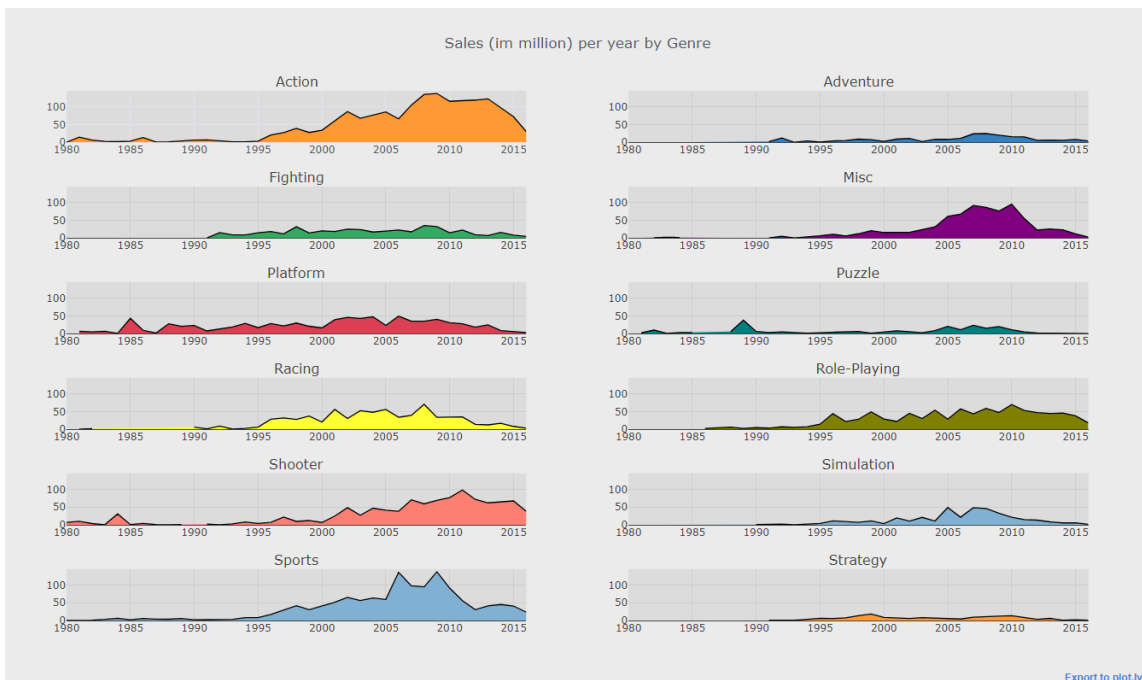
#### 4. Have the popular genres changed over time?

For this question, we obviously wanted to have a line chart, but showing all the lines on the same graph wasn't a good idea:



With too many genres (12), each with its own trend, this proved to be too complex as a visual aid- the colors end up too similar and it's difficult to follow one line and compare it to the others.

So instead, we chose to use small multiples:



While it is harder to compare the trends (since they're not all on the same plane) it did show individual trends much better, for example- the rise in the popularity of the Action and Shooter genres.

#### 4. Evaluation:

Pros:

- Area graphs are easy to read, can compare between the trends fairly well
- Different colors are more visually engaging

Cons:

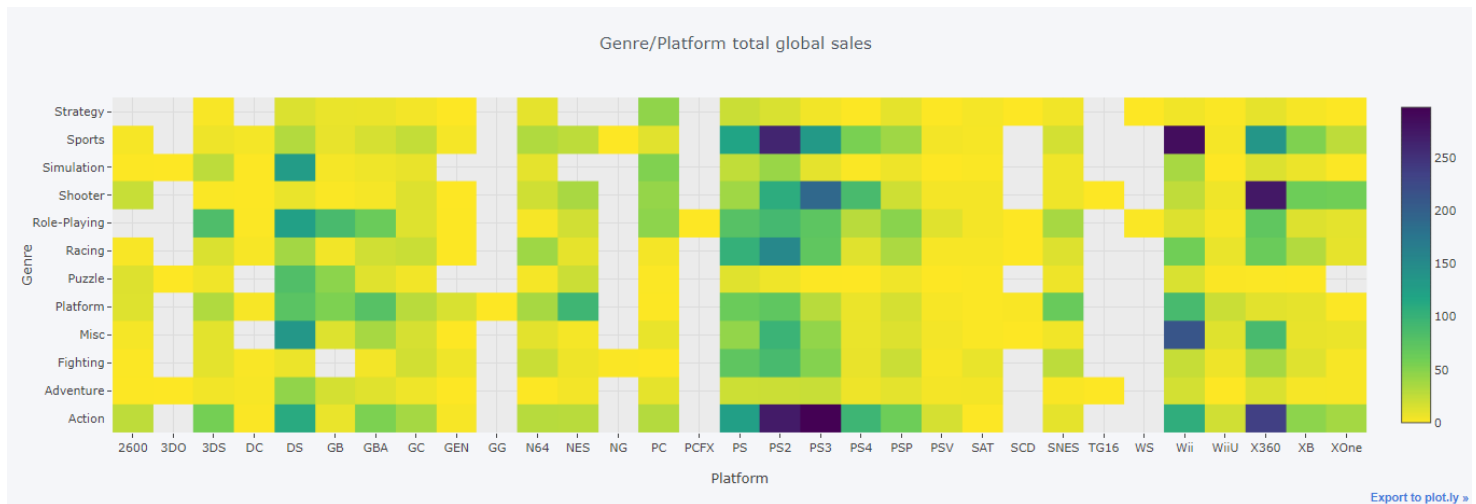
- Can't compare trends which are too similar (e.g. Strategy and Adventure)
- Keeping the same Unit of Measurement between all the graphs makes it difficult to see the trends in the lower values (e.g. Adventure vs Action)

$$V = T(20) + I(20) + E(25) + C(25) = 90\%$$

#### 5. Is there any correlation between Genre and Platform? Has this changed over time?

##### a. Is there any correlation between Genre and Platform?

To answer this, we've created a heat map, with the x-axis as the platform, and the y-axis as the genre:



While there are many holes in the data (caused by obscure platforms which supported only a few games)

It does provide the ability to see which genres are more successful in which platforms.

For example- Puzzle games were more successful on the DS and GameBoy consoles.

It also shows quite clearly the popular platforms: PlayStations1/2/3, Wii, Xbox360 and the DS.

## 5.a Evaluation:

Pros:

- Can find a specific platform/genre combination easily and check its value.

Cons:

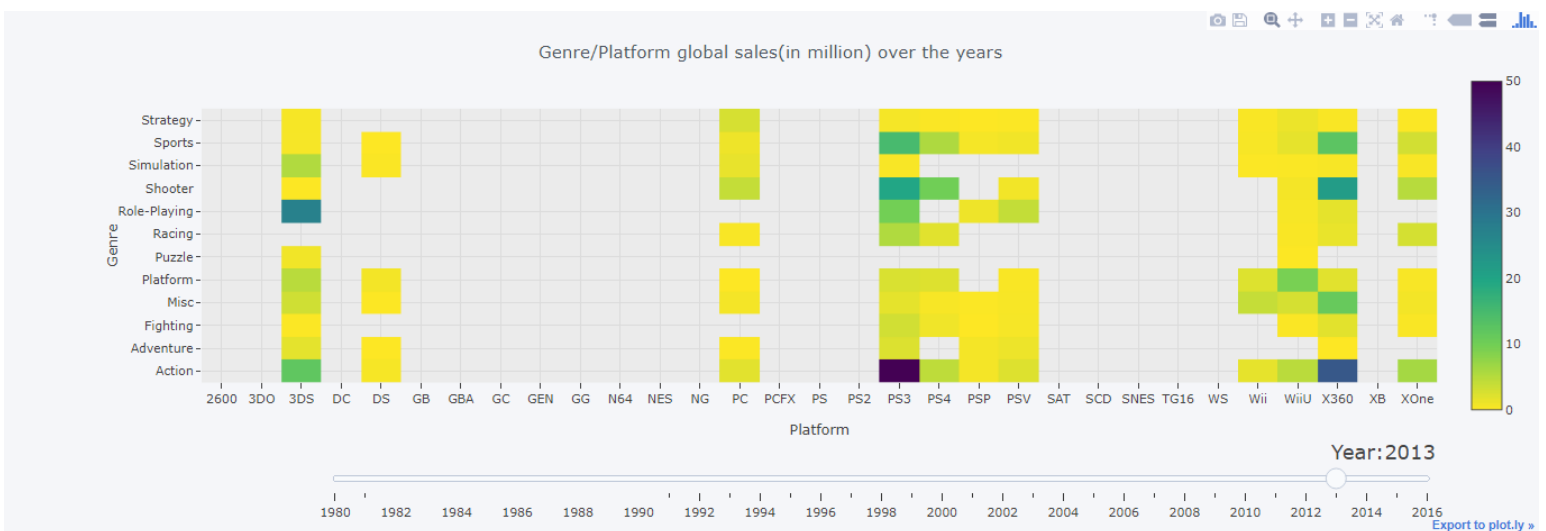
- Non sequential axes make it difficult to see trends or clusters, each square show an individual value and is not related to its neighbors.

$$V = T(20) + I(10) + E(15) + C(15) = 60\%$$

### b. Has it changed over time?

To answer how this changed over time though, we needed to show another axis.

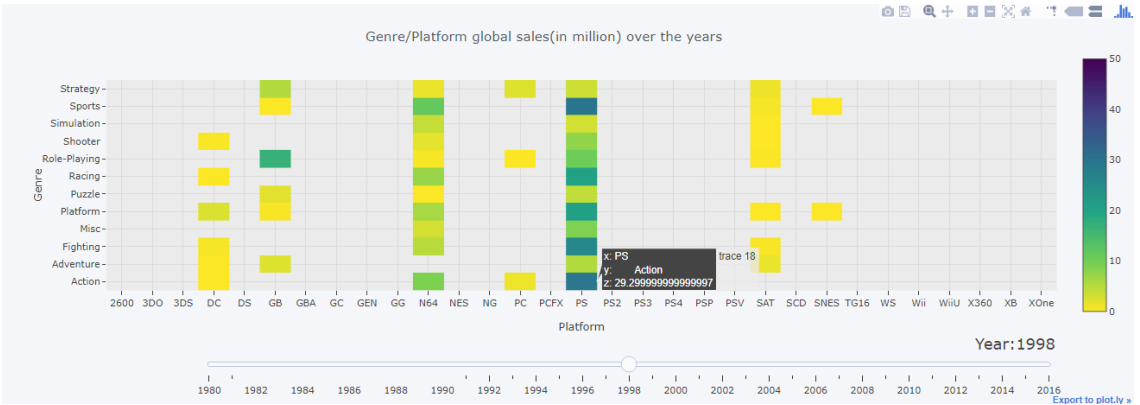
We chose to do this with an interactive slider:



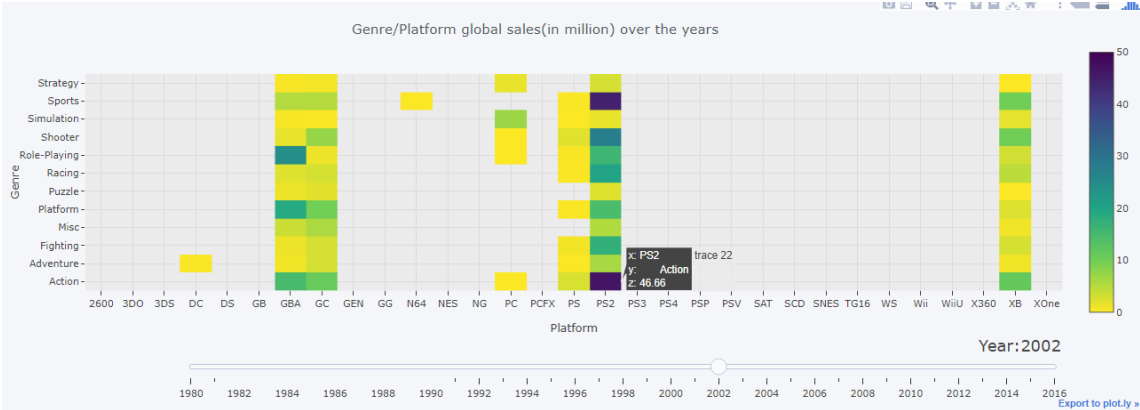
By sliding through the years and focusing on a single row, it was possible to see how the changes in preferred platform occur.

For example, we can see in Action games how the preferred platform updates through the years as newer versions of PlayStation come out-

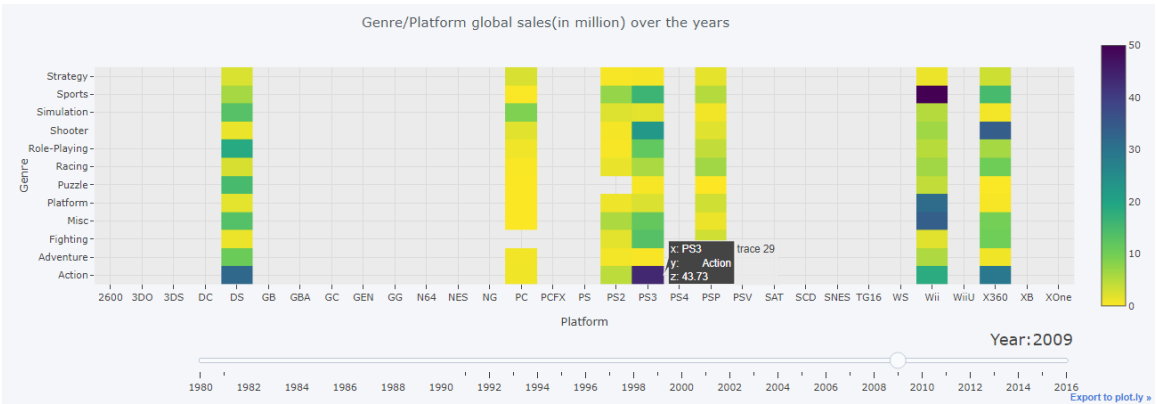
From 1998, where PlayStation 1 was at its highest:



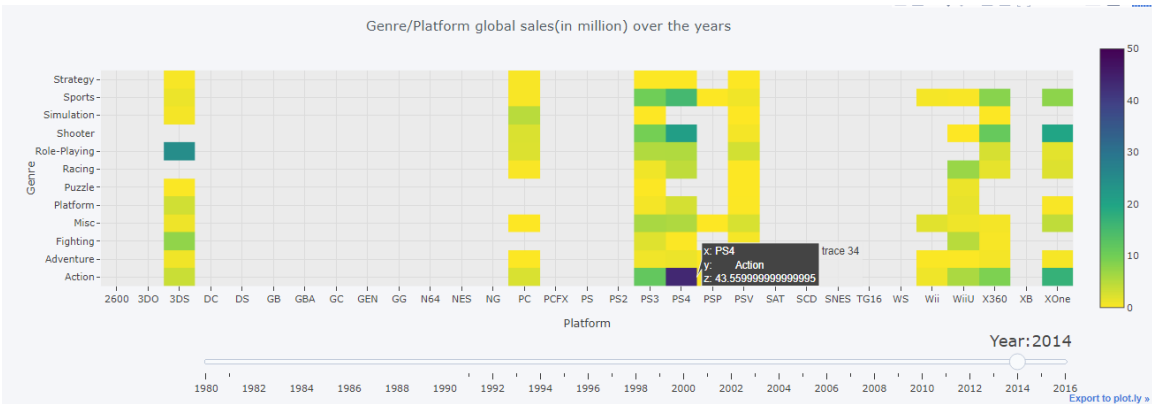
To 2002, where PlayStation2 was dominant:



To 2009 with PlayStation 3:



To 2014, with PlayStation 4



This graph provided a lot of information, but that information was not shown simply- you had to look at a specific row or column and keep track of it while sliding to gather any information from it.

However, since we're displaying a lot of information, we were unable to find a better way to show the correlation over time. With the exception of PlayStation the platforms themselves are not sequential, and so this was a difficult task.

### 5.b Evaluation:

Pros:

- With the addition of the slider it's possible to view the previous information per year, making looking up a specific value in our data set easier.

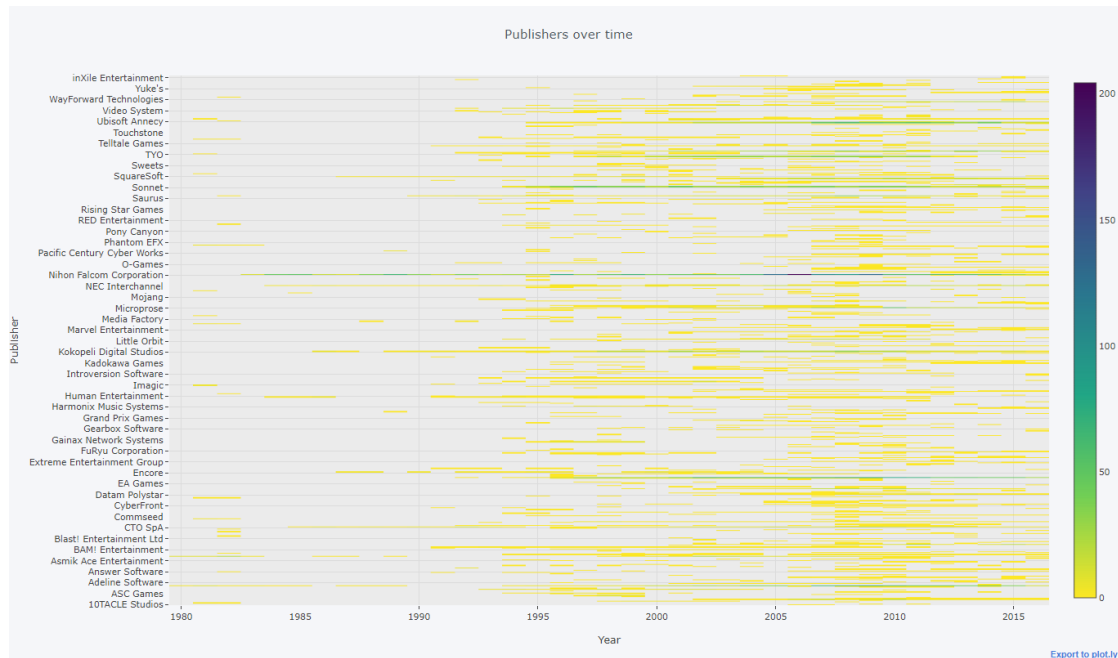
Cons:

- Non sequential axes make it difficult to see trends or clusters, each square show individual value and is not related to its neighbors.

$$V = T(20) + I(10) + E(15) + C(15) = 60\%$$

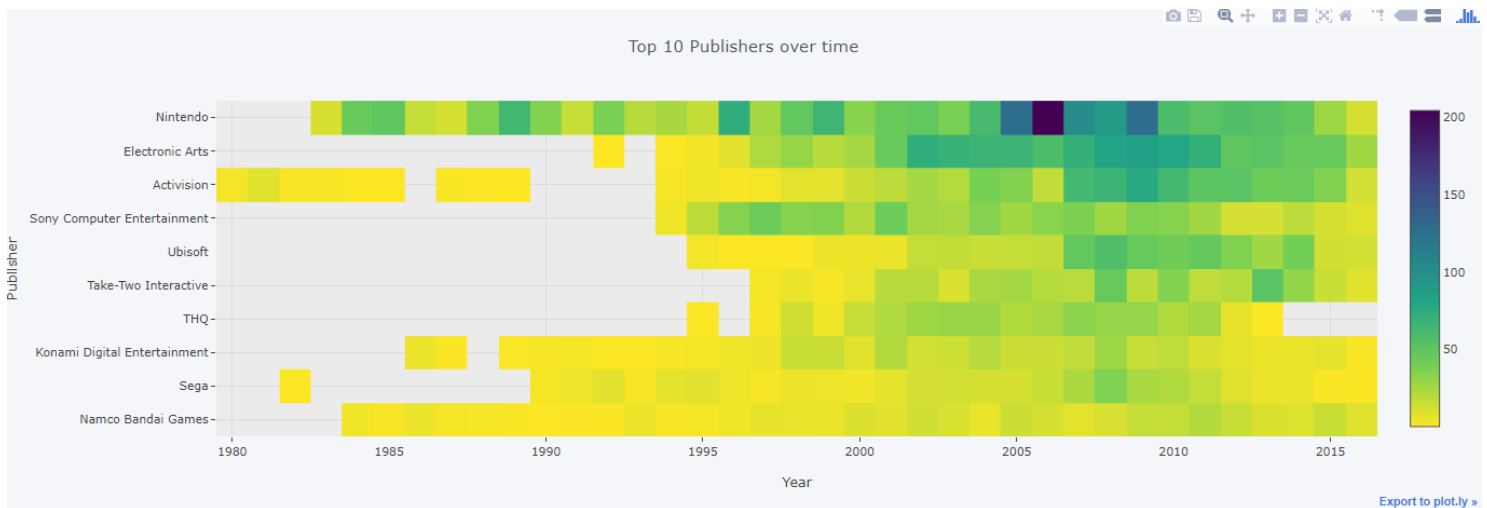
## 6. How did the publishers' sales change throughout the years?

We wanted to see how the publishers changed throughout the years.  
Originally we made a very ambitious heat map:



While it was possible to zoom in, this showed us nothing. There were too many publishers, most of them unknown.

So we chose to focus on the top 10 publishers, sorted from most successful (highest sales) to least successful:



This was more interesting- it shows the Nintendo Wii Sports spike in 2006, as well as how the top companies' popularity changed over the years.

- Nintendo shows a strong presence from the very start- consistently having high sales.
- Ubisoft and Electronic Arts and Activision slowly gain popularity over time.
- Sony losing popularity over time

## 6. Evaluation:

### Pros:

- Focusing only on the top 10 developers made it easier to see trends in their individual sales
- The sorting of the y-axis made it easier to compare the top publishers

### Cons:

- Sensitivity to outliers- Wii Sports pushes our scale to its max value, leaving the rest of the data to be more yellow, while without the Wii-Sports we would be able to see more variation.

$$V = T(20) + I(25) + E(20) + C(25) = 90\%$$



# Summary

We tried our hand at different graphs, trying to find and visualize the most interesting points in the data.

Our data was rather extensive, with a lot of different aspects, which made finding a single graph to represent all of it quite challenging.

We've learned that sometimes it is best to focus on smaller, more relevant areas of the data instead of trying to display a full- more complex image, as can be seen in the publishers graph for example (choosing to focus on the interesting, top 10, as opposed to all of them at once).

We've also learned that sometimes it's better to have multiple graphs, instead of 1 busy graph, as was the case with the genre trends (small multiples)

Eventually we believe we came up with some interesting approaches to our questions.

With each graph there were many choices to consider, trying to see what would best show the data, which was a rather difficult task, but also an interesting one.