Data Wrangling and Husbandry

Analyzing Video Games Sales

Github Link: https://github.com/96malhar/Data-Wrangling-Final-Project

Abstract

A video game is an electronic game that can be played on a computing device, such as a personal computer, gaming console or mobile phone. Depending on the platform, video games can be subcategorized into computer games and console games. In recent years however, the emergence of social networks, smartphones and tablets introduced new categories such as mobile and social games. Video games have come a long way since the first games emerged in the 1970s. Today's video games offer photorealistic graphics and simulate reality to a degree which is astonishing in many cases.

2018 was a record-breaking year for our industry, with total video game sales exceeding $43.4 billion. Over 164 million adults in the United States play video games and three-quarters of all Americans have at least one gamer in their household. As the leading form of entertainment today, video games are an integral part of American culture. To that end, I decided to perform some analysis in order to understand what is driving the video game industry.

Data Schema

The Data is a CSV file that was scraped from https://www.vgchartz.com/. It contains the following columns.

- **Rank** - Ranking of overall sales
- **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.

Libraries Used
- **ggplot2**
    - ggplot2 is a system for declaratively creating graphics, based on The Grammar of Graphics. You provide the data, tell ggplot2 how to map variables to aesthetics, what graphical primitives to use, and it takes care of the details.
- **Tidyr**

- o Tidy data describes a standard way of storing data that is used wherever possible throughout the tidyverse. If you ensure that your data is tidy, you'll spend less time fighting with the tools and more time working on your analysis.
- **DT**
  - o The R package DT provides an R interface to the JavaScript library DataTables. R data objects (matrices or data frames) can be displayed as tables on HTML pages, and DataTables provides filtering, pagination, sorting, and many other features in the tables.
- **dplyr**
  - o dplyr is a grammar of data manipulation, providing a consistent set of verbs that help you solve the most common data manipulation challenges:
    - mutate() adds new variables that are functions of existing variables
    - select() picks variables based on their names.
    - filter() picks cases based on their values.
    - summarise() reduces multiple values down to a single summary.
    - arrange() changes the ordering of the rows.

Analysis



Video Game Releases by Year
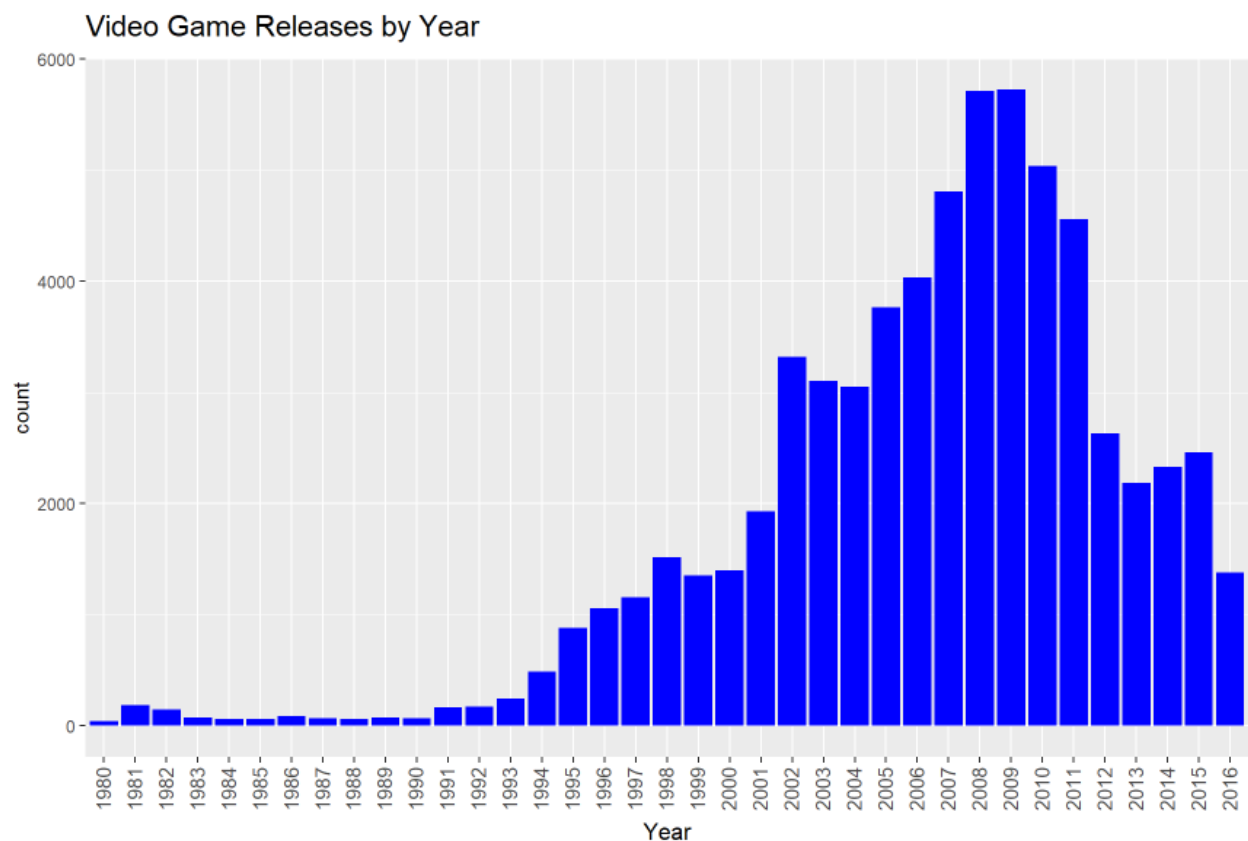
Video Game Revenue by Year
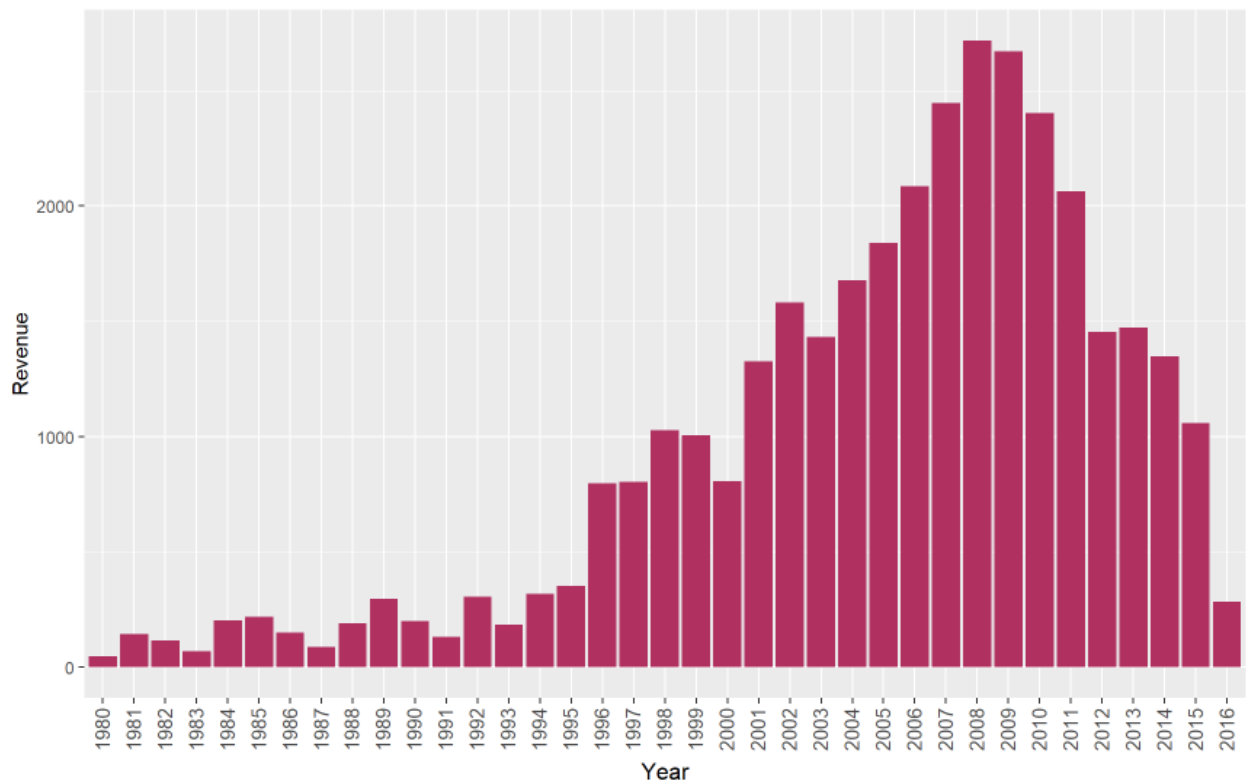
There is huge spike in the number of releases after 2000 and it peaked during 2008 and 2009. It has started to decline in the last few years. This is surprising as higher demand should lead to greater revenue

Top Publisher by Revenue each Year

Publisher ▮ Activision ▮ Atari ▮ Electronic Arts ▮ Nintendo ▮ Sony Computer Entertainment
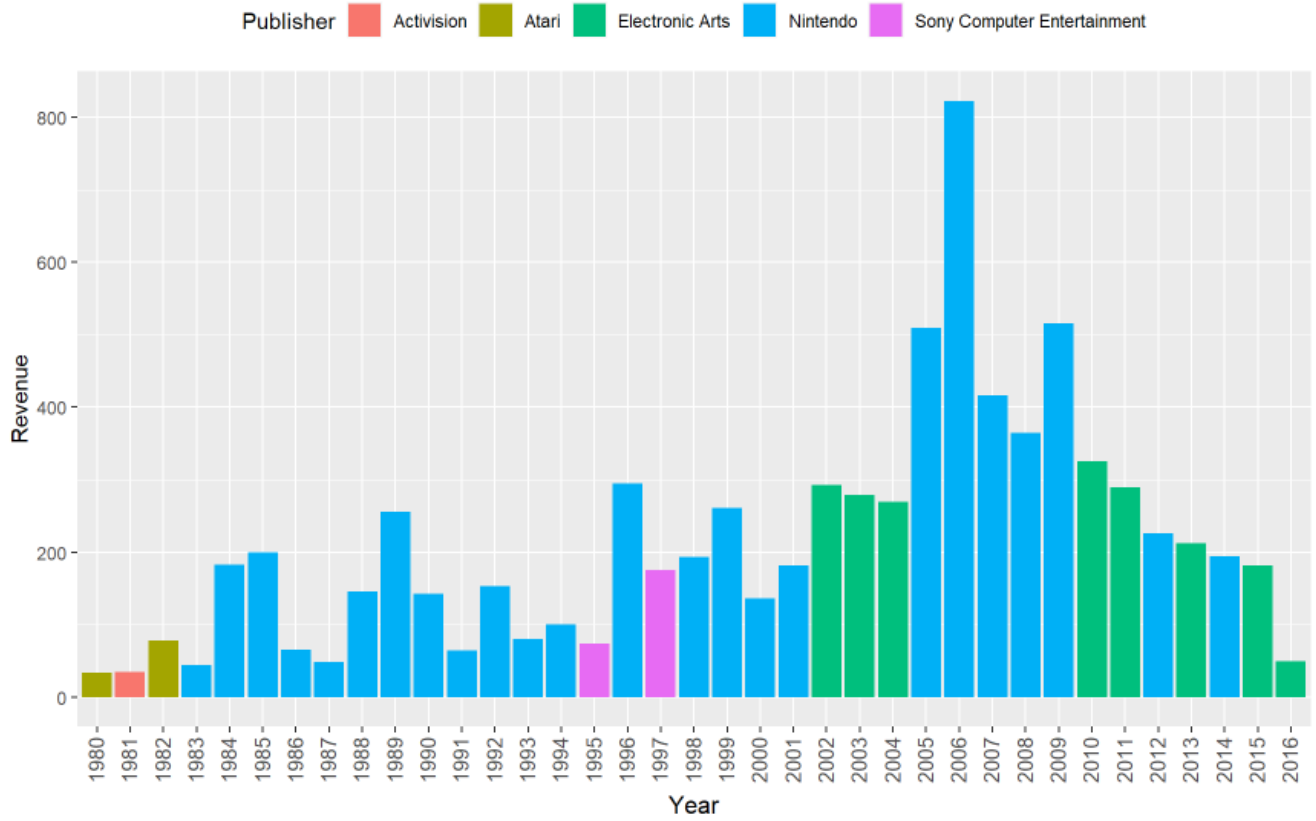
Nintendo is by far the most dominant company in the video game industry. Along with Nintendo, Electronic Arts is also dominating the market since the last decade.

Top Genre by Revenue each Year
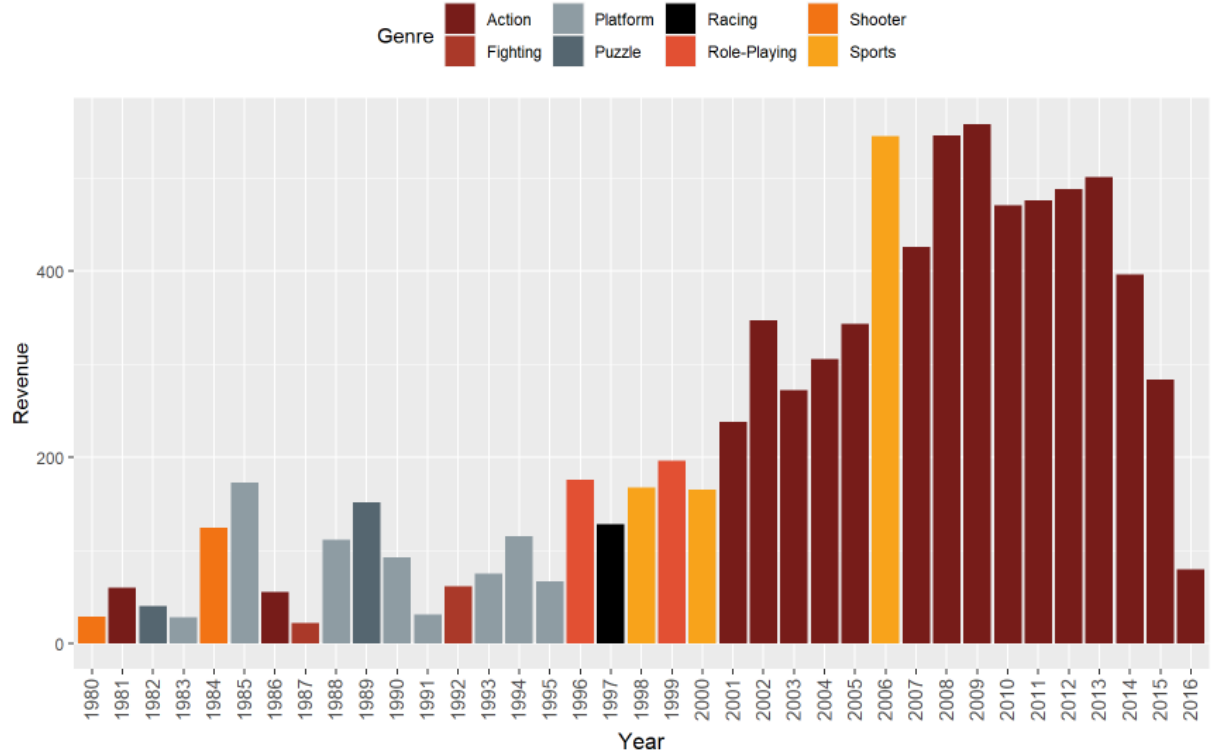
Action has been the highest revenue generating genre in the last 15 years, except for once in 2006, during which Sports was the highest revenue generator.

Top Platform by Revenue each year

Platform: ▇ 2600 ▇ NES ▇ PS2 ▇ PS4 ▇ Wii
▇ GB ▇ PS ▇ PS3 ▇ SNES ▇ X360
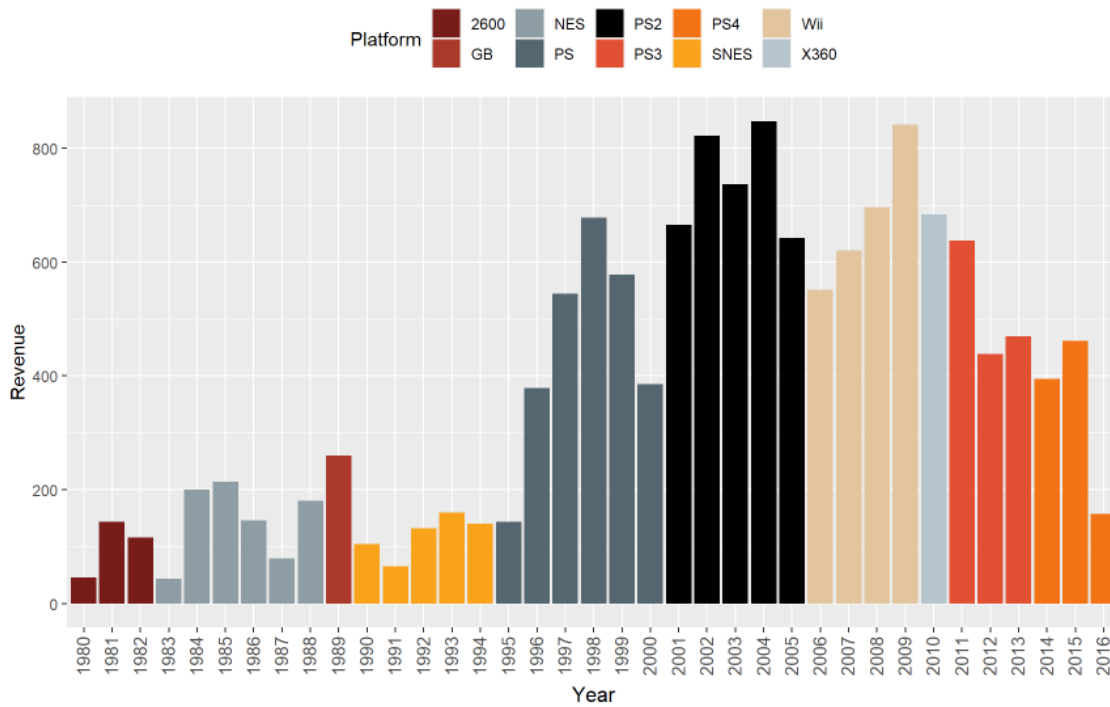
Gaming platforms developed by Nintendo were popular in in the 1980s and early 1990s. However, Playstation has been the most popular console since 1995 with PS1 dominating the market since 1995 followed by PS2 in 2001. Nintendo rose back to popularity with its Wii console in 2006 only to be replaced by subsequent generation of the Playstation family in 2011.

Top 10 Publishers by Number of Releases
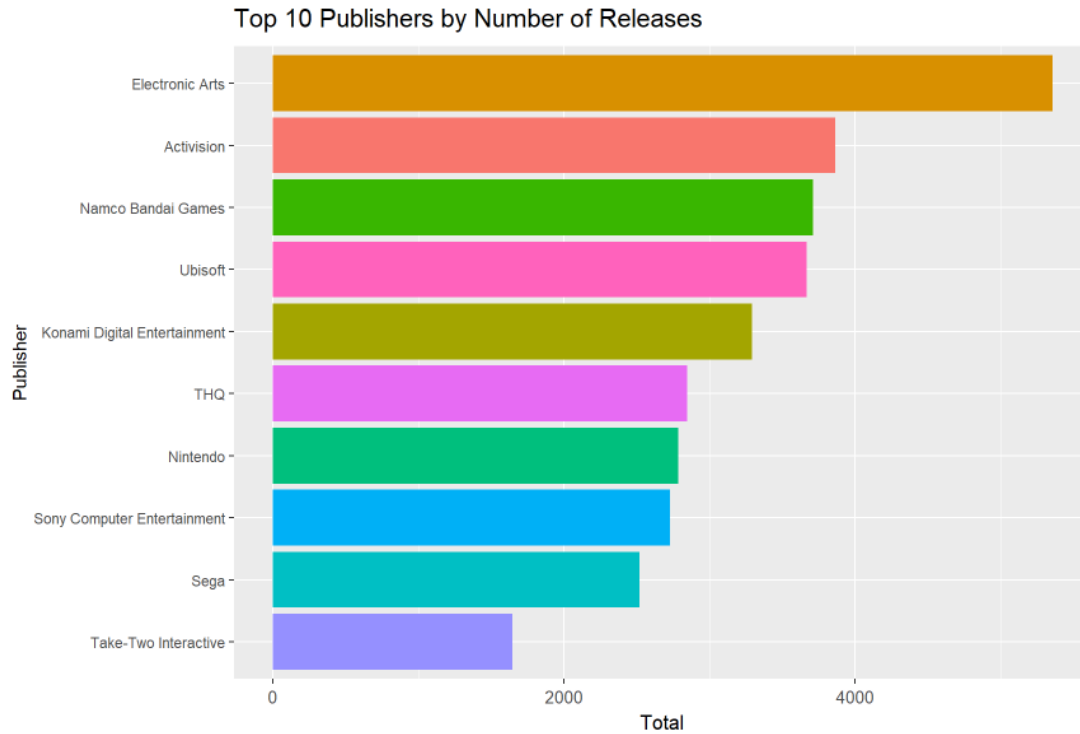
There are 577 publishers with Electronic Arts being the top publisher followed by Activision. Top 10 publishers have published nearly 35% of all the games.

Top 10 Publishers by Revenue
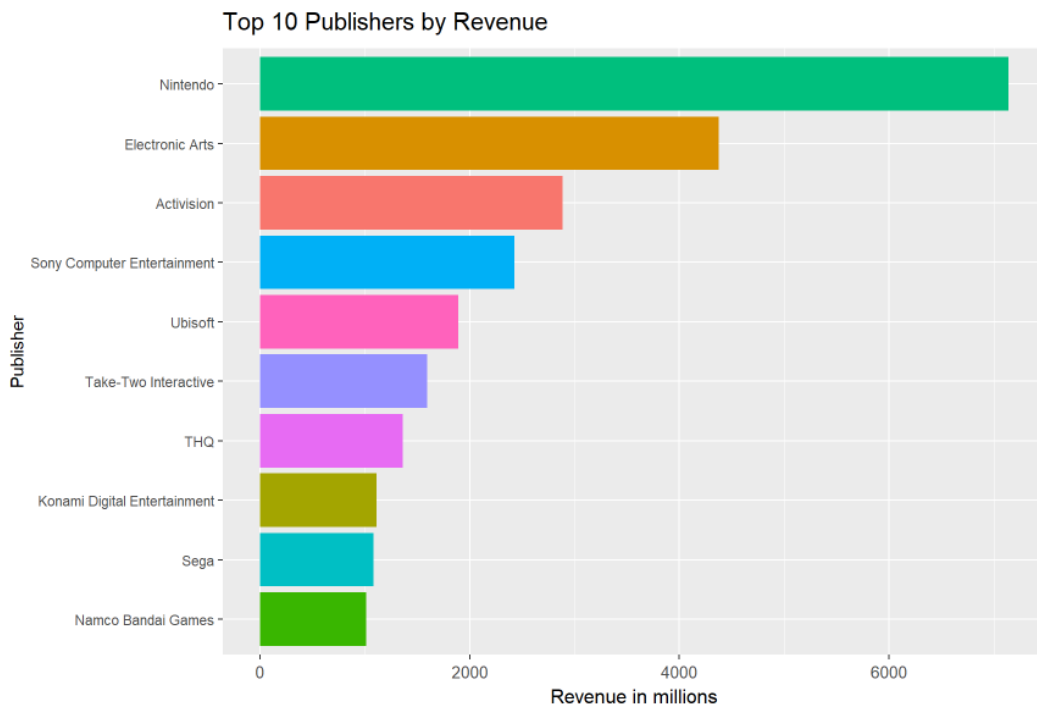
As shown by the above plot, more number of releases does not imply higher revenue. Despite Nintendo ranking 6th in terms of number of release with its numbers being half of that of Electronic Arts, Nintendo ranks 1st in terms of the Revenue generated. Electronic Arts comes 2nd in terms of revenue generated.


Top 10 Publishers Releases by Year
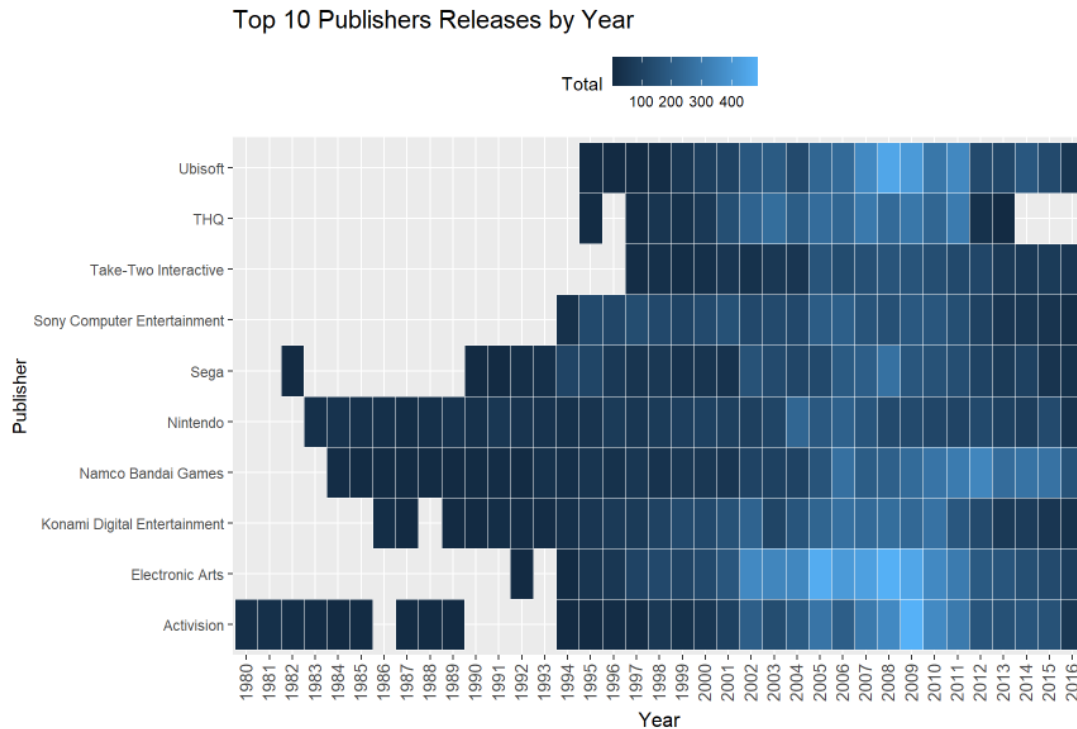
Electronic Arts did not release its first game until 1992. Activision along with Sega have been the oldest publishers in this industry. EA has highest number of releases between 2002 and 2011 and THQ has not released any games from 2014

## Video Games Genres by Number of Releases
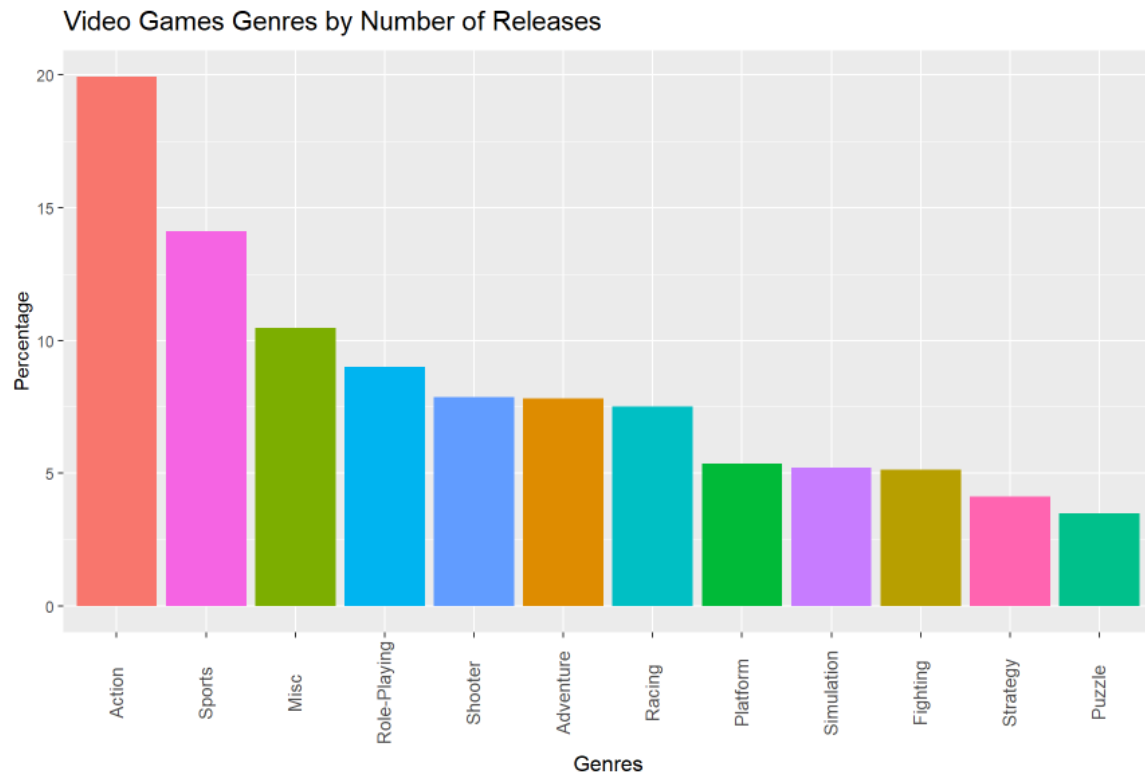
Action Games contribute almost 20% of the games released with Sports games coming in second. The top 5 genres contribute to 60% of the games released.

## Video Games Genres by Revenue
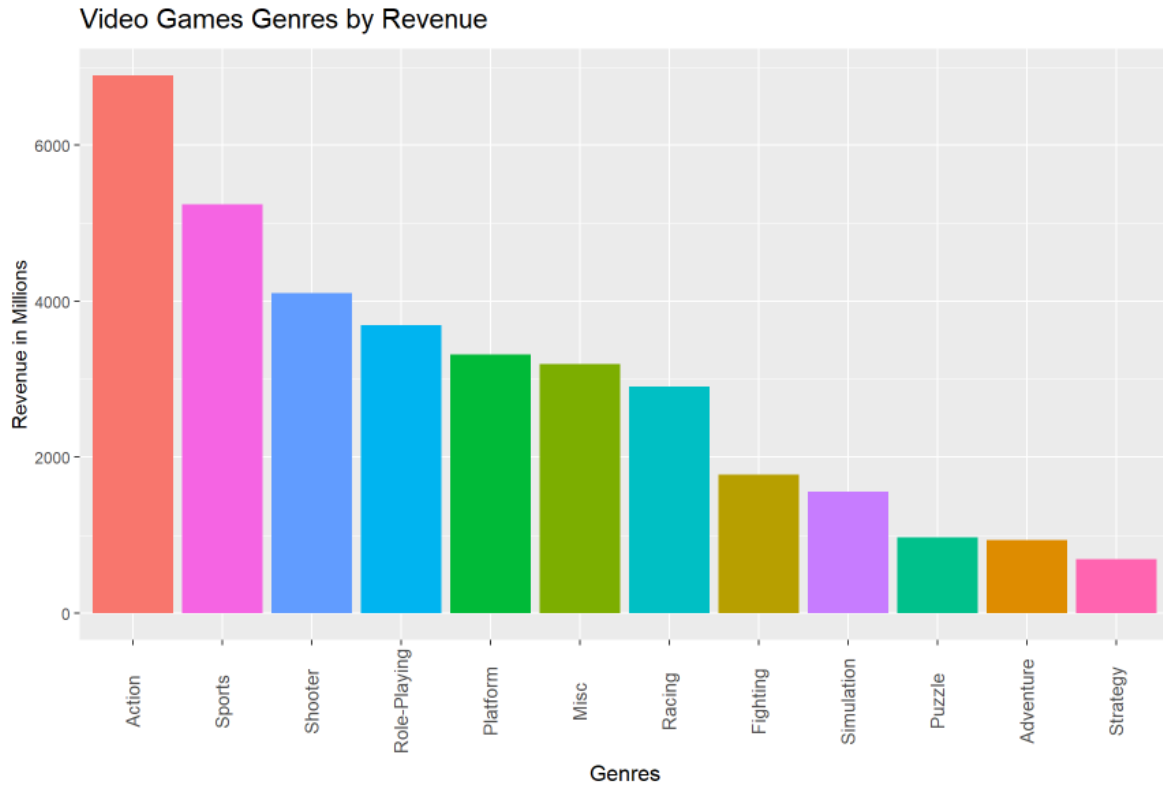
Action Games contribute to almost 20% of the revenue. Sports games coming in second with 15% of the revenue. Overall all top 5 genres contribute to 65% of the revenue.

Video Games Genre Ranking based on Revenue Per Release
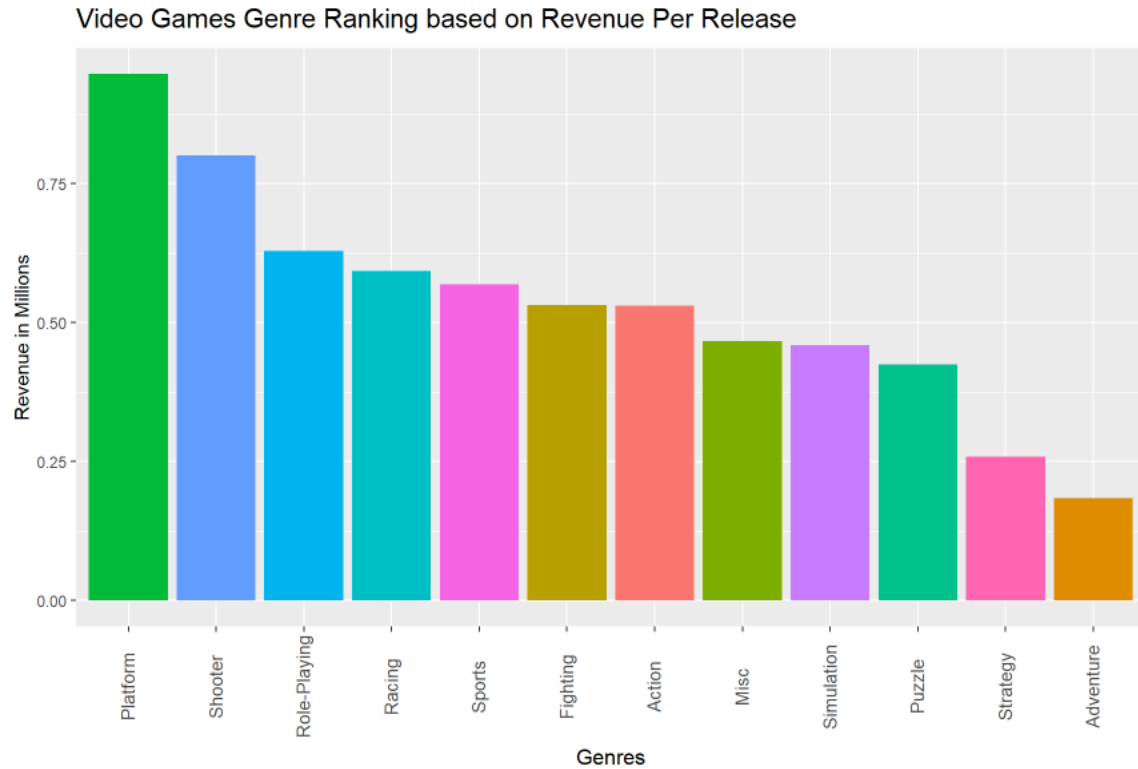
Platform is the best genre by revenue per release. Shooter and Role Playing with second and third best.

Top 10 Games by Revenue
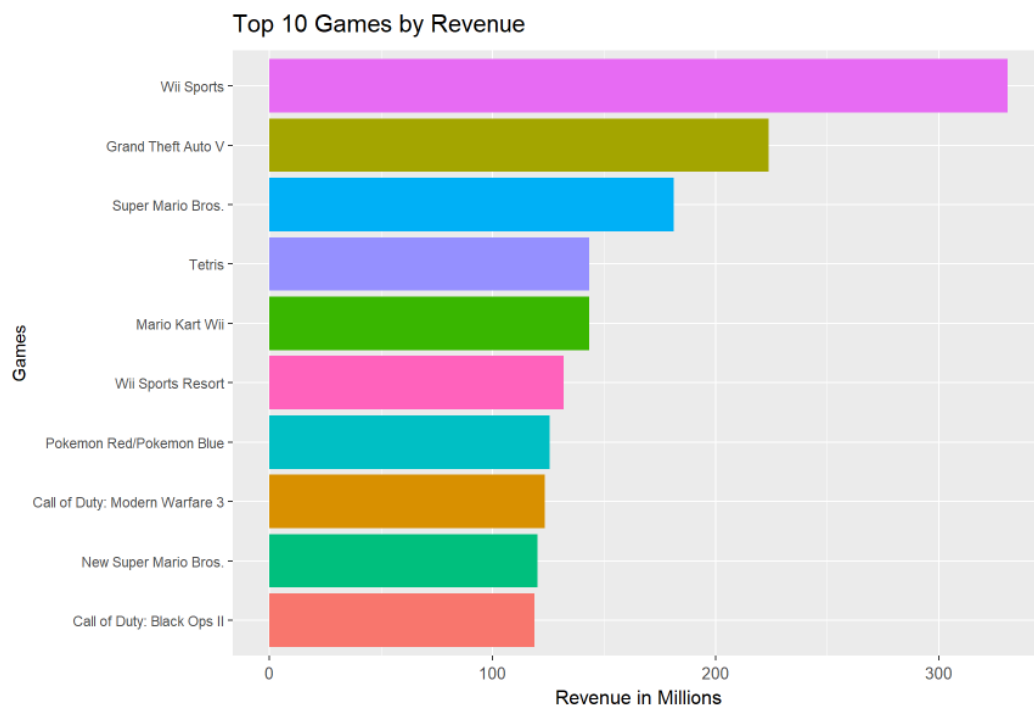
Wii Sports game has 1% of the total revenue with 82.74 million. Grand Theft Auto V coming second with 55.92 million

Total Revenue by Region
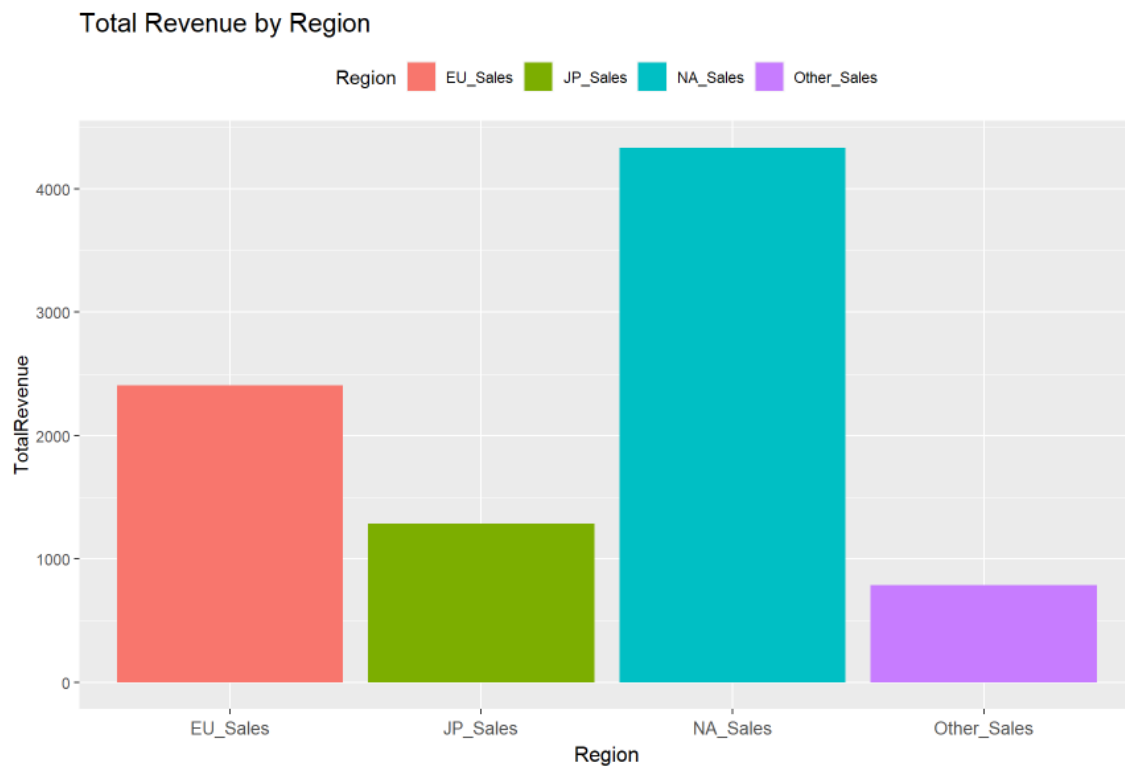
North America market leads the revenue rankings as it contributes 49% of the revenue. Europe contributes 27% of the revenue.

Distribution of Sales Revenue
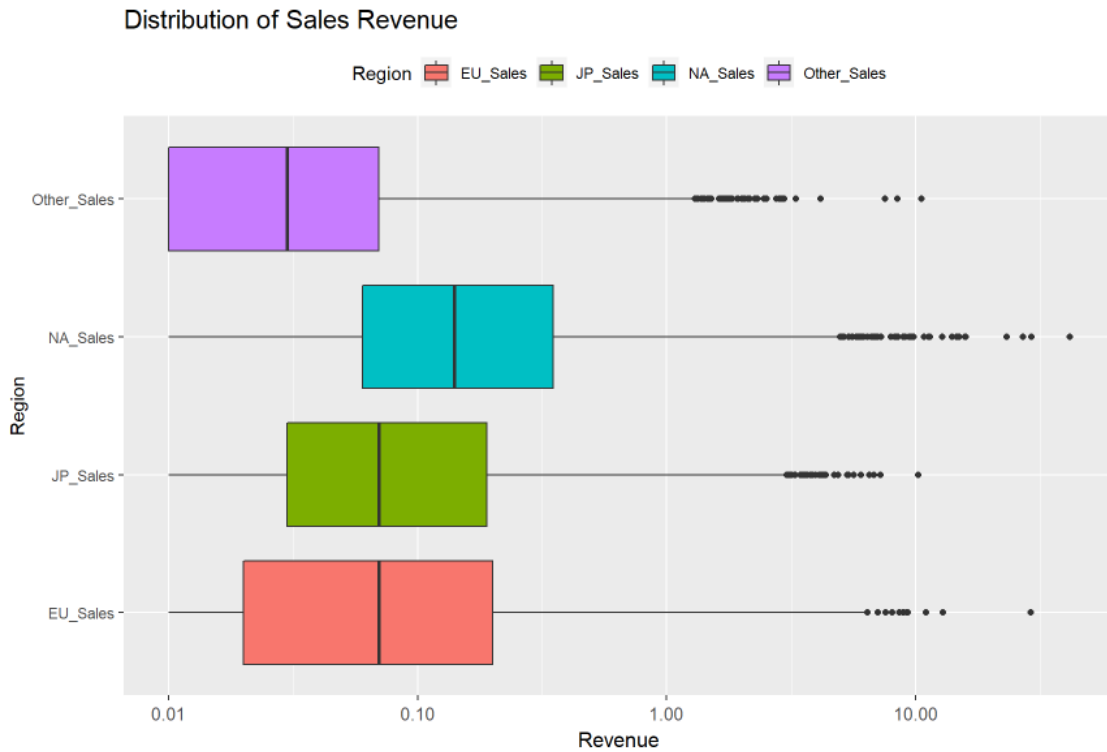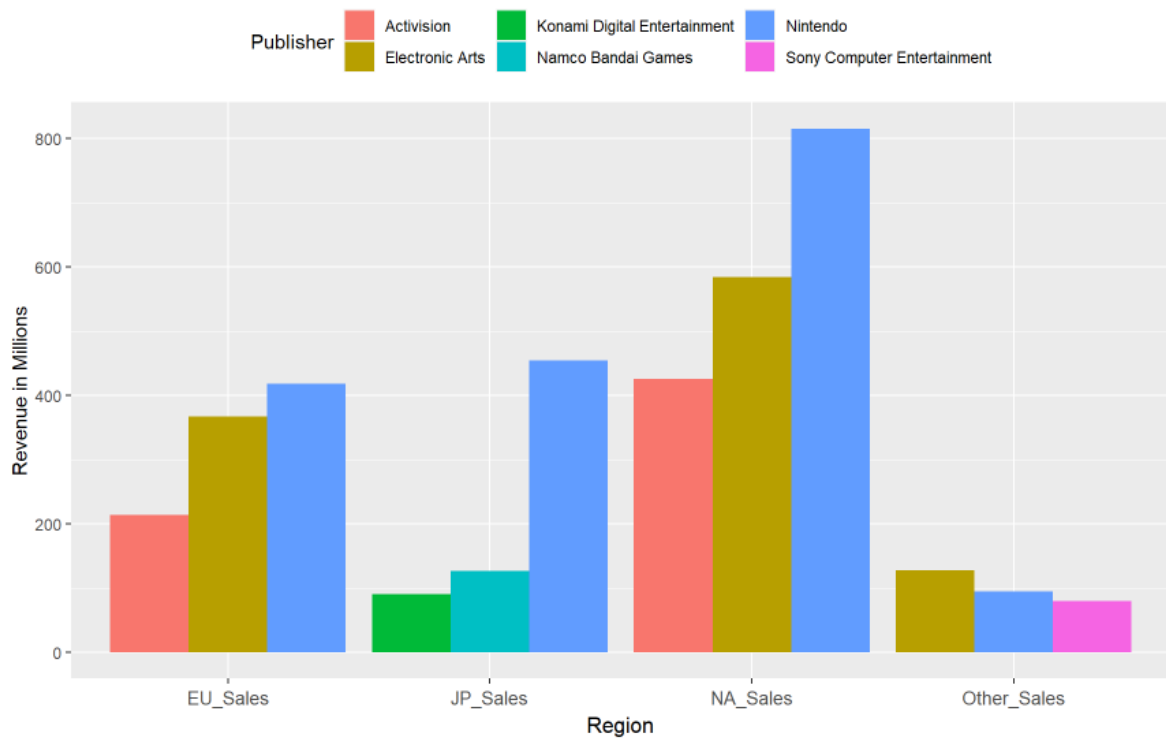

North America has the highest median revenue. Europe and Japan are almost equal in terms of median revenue.

Total Revenue by Region and Year
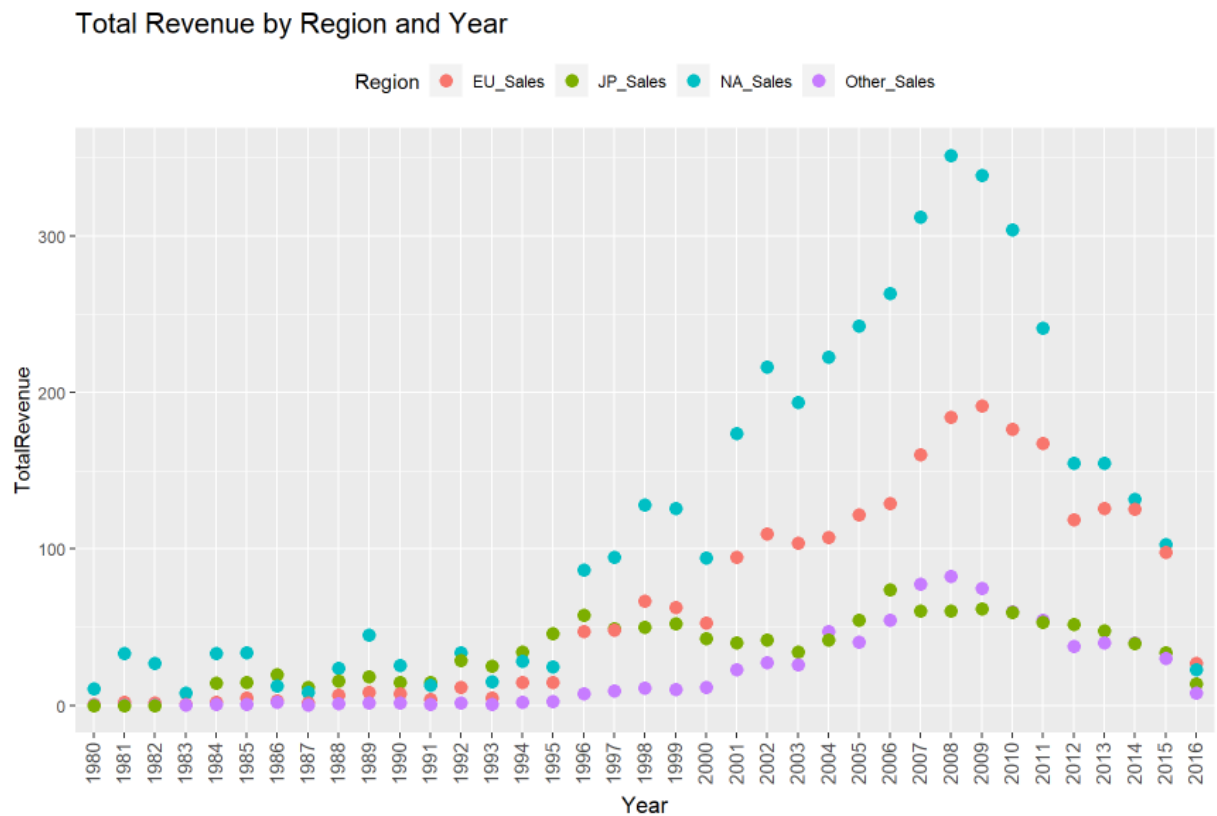
Total Revenue in North America has been consistently higher than other regions. Total revenue has an increasing trend till 2010 but has been steadily decreasing in the last decade.
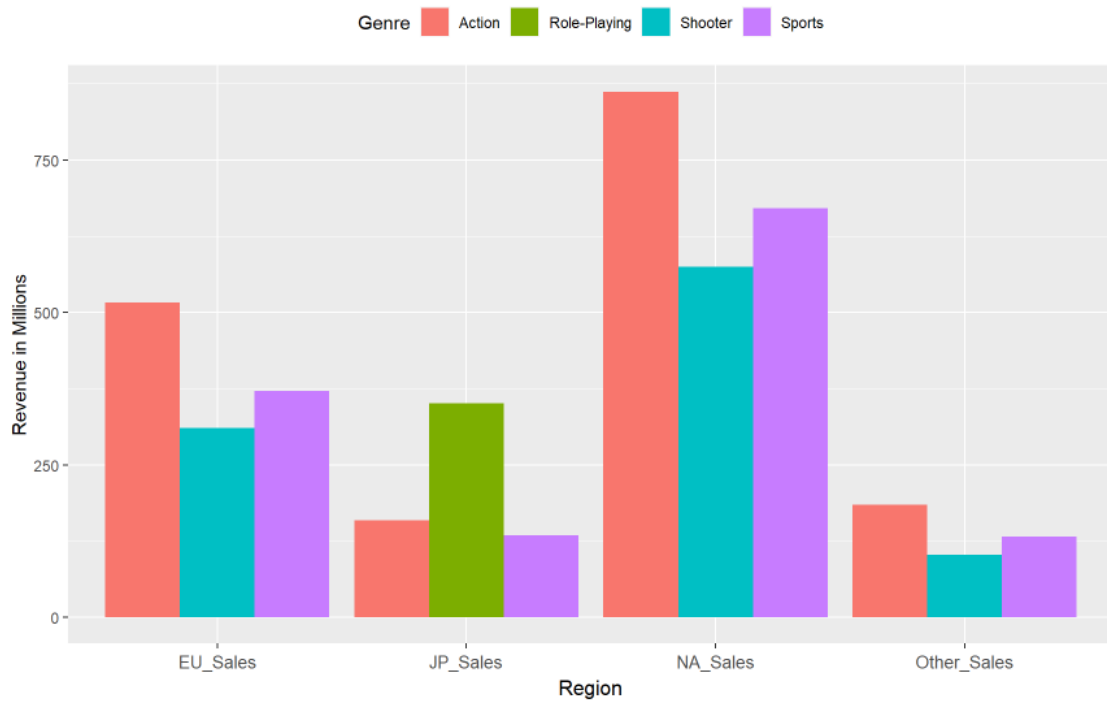
Top 3 Publishers by Revenue in each Region

Nintendo being a Japanese company completely dominates the market in Japan. Electronic Arts is also very popular in North America as they develop sports simulation Games Like FIFA and NBA. Both Soccer and Basket Ball are extremely popular in Europe and North America respectively.

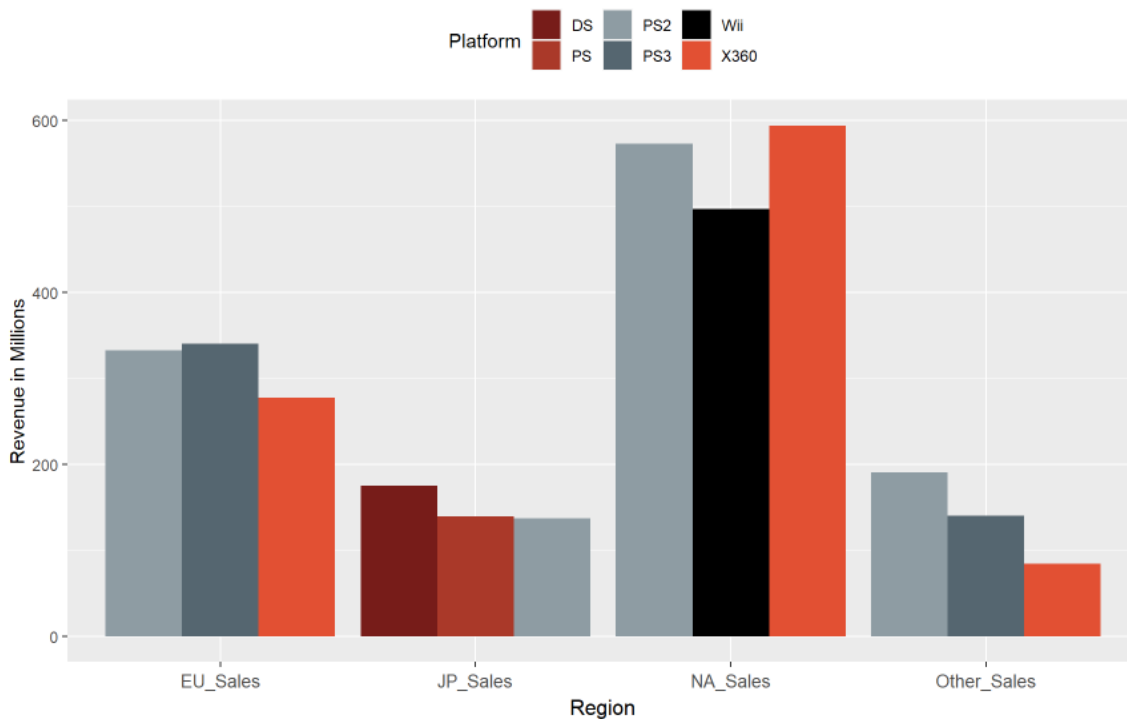Top 3 Games by Revenue in each Region
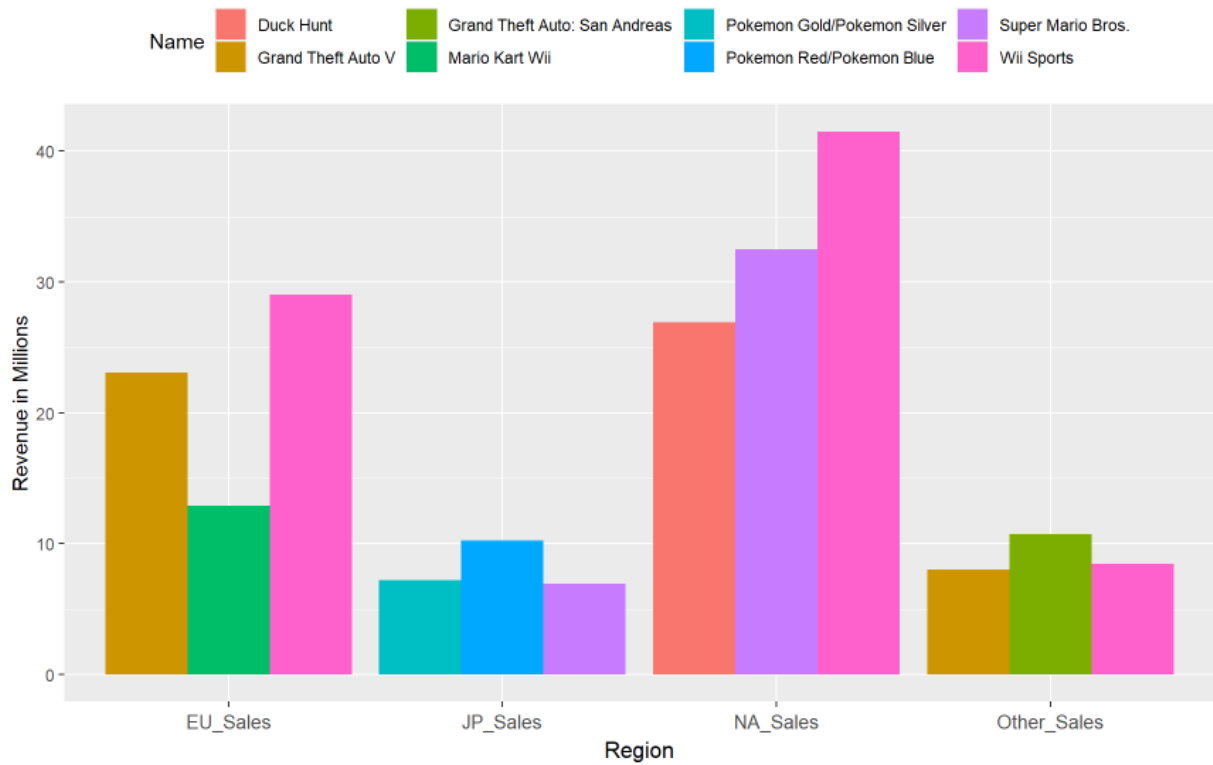
Top 3 Genres by Revenue in each Region

Action is the most popular Genre across all regions.

Top 3 Platform by Revenue in each Region

Xbox is developed by Microsoft which is an American Company and has the highest revenue in North America. Similarly DS is developed by Nintendo which is a Japanese company and generates the highest revenue in Japan.

Conclusion

- o Over the last few years there is decrease in the games which have crossed 100K in sales.
- o last 5-6 years has seen a decrease in the overall revenue generated.
- o Nintendo generates the most revenue across all publishers.
- o Action generates the most revenue across all genres.
- o Action is also the most popular genre in the market since 2001
- o Play Station is the most popular platform
- o Electronic Arts ranks 1st in terms of number of release number of releases.
- o Platform Genre generated most revenue per release.
- o Wii Sports is ranks 1st in terms of revenue.
- o North America and Europe account for 75% of the market share by revenue.

References

- o https://www.vgchartz.com/
- o https://towardsdatascience.com/analyzing-video-games-data-in-r-1afad7122aab
- o https://medium.com/video-games-sales/exploratory-data-analysis-e634689e3a52

- https://tidyr.tidyverse.org/
- https://dplyr.tidyverse.org/
- https://ggplot2.tidyverse.org/