

# Video Games Sales Analysis

Pratik Hotchandani

2023-02-02

```
library(dplyr)

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union

#Library to import the data
library(readr)

# This dataset on Video games sales is downloaded from Kaggle.
# Dataset link - https://www.kaggle.com/datasets/rush4ratio/video-game-sales-with-ratings

df <- read_csv("Video_Games_Sales_as_at_22_Dec_2016.csv")
```

This dataset is prepared by web scraping the website 'Metacritic'. This dataset is actually a combination of two datasets, one had the sales related information of the games and the second dataset has included game ratings by users and critics.

These are the variables included:

1. Name: Title of the video game
2. Platform: Video gaming console on which the game is released.
3. Year\_of\_Release: The year when the video game was released.
4. Genre: The genre of the game.
5. Publisher: The publisher who published the game.
6. Developer: The Developer who developed the game.
7. NA\_Sales: Sales in North America.
8. EU\_Sales: Sales in Europe.

9. JP\_Sales: Sales in Japan.
10. Other\_Sales: Sales in other countries.
11. Global\_Sales: Overall sales of the title.
12. Critic\_Score: Aggregate score compiled by Metacritic staff.
13. Critic\_Count: The number of critics used in coming with the critic score.
14. User\_Score: Score by Metacritics subscribers.
15. Rating: The ESRB rating of the game.
16. User\_Count: Number of users who gave the user\_score.

```
# summary of the dataset
summary(df)
```

```
##      Name      Platform      Year_of_Release      Genre
## Length:16719 Length:16719 Length:16719 Length:16719
## Class :character Class :character Class :character Class :character
## Mode  :character Mode  :character Mode  :character Mode  :character
##
##
##
##
##      Publisher      NA_Sales      EU_Sales      JP_Sales
## Length:16719 Min.   : 0.0000 Min.   : 0.000 Min.   : 0.0000
## Class :character 1st Qu.: 0.0000 1st Qu.: 0.000 1st Qu.: 0.0000
## Mode  :character Median : 0.0800 Median : 0.020 Median : 0.0000
##                  Mean   : 0.2633 Mean   : 0.145 Mean   : 0.0776
##                  3rd Qu.: 0.2400 3rd Qu.: 0.110 3rd Qu.: 0.0400
##                  Max.   :41.3600 Max.   :28.960 Max.   :10.2200
##
##      Other_Sales      Global_Sales      Critic_Score      Critic_Count
## Min.   : 0.00000 Min.   : 0.0100 Min.   :13.00 Min.   : 3.00
## 1st Qu.: 0.00000 1st Qu.: 0.0600 1st Qu.:60.00 1st Qu.: 12.00
## Median : 0.01000 Median : 0.1700 Median :71.00 Median : 21.00
## Mean   : 0.04733 Mean   : 0.5335 Mean   :68.97 Mean   : 26.36
## 3rd Qu.: 0.03000 3rd Qu.: 0.4700 3rd Qu.:79.00 3rd Qu.: 36.00
## Max.   :10.57000 Max.   :82.5300 Max.   :98.00 Max.   :113.00
##                  NA's   :8582 NA's   :8582
##      User_Score      User_Count      Developer      Rating
## Length:16719 Min.   : 4.0 Length:16719 Length:16719
## Class :character 1st Qu.: 10.0 Class :character Class :character
## Mode  :character Median : 24.0 Mode  :character Mode  :character
##                  Mean   : 162.2
##                  3rd Qu.: 81.0
##                  Max.   :10665.0
##                  NA's   :9129
```

### *Data pre processing*

1. From the summary created, we can see that there are NA values in few columns. Removing the null values will help us understand the data better.

```
df_cleaned <- na.omit(df)
df_cleaned
```

```
## # A tibble: 6,947 x 16
##   Name      Platf~1 Year_~2 Genre Publi~3 NA_Sa~4 EU_Sa~5 JP_Sa~6 Other~7 Globa~8
##   <chr>    <chr>    <chr>    <chr> <chr>    <dbl>    <dbl>    <dbl>    <dbl>    <dbl>
## 1 Wii Sp~ Wii      2006    Spor~ Ninten~  41.4    29.0     3.77    8.45    82.5
## 2 Mario ~ Wii      2008    Raci~ Ninten~  15.7    12.8     3.79    3.29    35.5
## 3 Wii Sp~ Wii      2009    Spor~ Ninten~  15.6    10.9     3.28    2.95    32.8
## 4 New Su~ DS        2006    Plat~ Ninten~  11.3     9.14    6.5     2.88    29.8
## 5 Wii Pl~ Wii      2006    Misc  Ninten~  14.0     9.18    2.93    2.84    28.9
## 6 New Su~ Wii      2009    Plat~ Ninten~  14.4     6.94    4.7     2.24    28.3
## 7 Mario ~ DS        2005    Raci~ Ninten~   9.71    7.47    4.13    1.9     23.2
## 8 Wii Fit Wii      2007    Spor~ Ninten~   8.92    8.03    3.6     2.15    22.7
## 9 Kinect~ X360     2010    Misc  Micros~  15      4.89    0.24    1.69    21.8
## 10 Wii Fi~ Wii      2009    Spor~ Ninten~   9.01    8.49    2.53    1.77    21.8
## # ... with 6,937 more rows, 6 more variables: Critic_Score <dbl>,
## #   Critic_Count <dbl>, User_Score <chr>, User_Count <dbl>, Developer <chr>,
## #   Rating <chr>, and abbreviated variable names 1: Platform,
## #   2: Year_of_Release, 3: Publisher, 4: NA_Sales, 5: EU_Sales, 6: JP_Sales,
## #   7: Other_Sales, 8: Global_Sales
```

---

2. Next we will check if each column has a correct class-type.

```
# Get the class of each column
sapply(df_cleaned, class)
```

```
##           Name           Platform Year_of_Release           Genre           Publisher
## "character" "character"    "character"    "character"    "character"
##      NA_Sales      EU_Sales      JP_Sales      Other_Sales      Global_Sales
## "numeric"    "numeric"    "numeric"    "numeric"    "numeric"
## Critic_Score Critic_Count User_Score      User_Count      Developer
## "numeric"    "numeric"    "character"    "numeric"    "character"
##           Rating
## "character"
```

---

3. We can see that the column 'Year of Release' and 'User\_Score' are character type, it should be numeric. Let's see all the data from column 'Year of Release'.

```
unique(df_cleaned$Year_of_Release)
```

```
## [1] "2006" "2008" "2009" "2005" "2007" "2010" "2013" "2004" "2002" "2001"
## [11] "2011" "2012" "2014" "1997" "1999" "2015" "2016" "2003" "1998" "1996"
## [21] "2000" "N/A"  "1994" "1985" "1992" "1988"
```

4. We can see that when removing the null values, “N/A” was missed since it is a character value. It is possible that for other character columns to have “N/A” values. Let us check for that.

```
sum(df_cleaned$Name=="N/A")
```

```
## [1] 0
```

```
sum(df_cleaned$Platform=="N/A")
```

```
## [1] 0
```

```
sum(df_cleaned$Genre=="N/A")
```

```
## [1] 0
```

```
sum(df_cleaned$Publisher=="N/A")
```

```
## [1] 4
```

```
sum(df_cleaned$User_Score=="N/A")
```

```
## [1] 0
```

```
sum(df_cleaned$Developer=="N/A")
```

```
## [1] 0
```

```
sum(df_cleaned$Rating=="N/A")
```

```
## [1] 0
```

```
sum(df_cleaned$Year_of_Release=="N/A")
```

```
## [1] 121
```

- 
5. From above results, we can see that “Year of Release” and “Publisher” columns has “N/A” value. Lets remove rows with N/A values.

```
df_cleaned <- df_cleaned[df_cleaned$Year_of_Release != "N/A",]  
df_cleaned <- df_cleaned[df_cleaned$Publisher != "N/A",]
```

- 
6. Since there are no more “N/A” values, we can now convert both the columns to integer type.

```
df_cleaned$Year_of_Release <- as.integer(df_cleaned$Year_of_Release)
df_cleaned$User_Score <- as.integer(df_cleaned$User_Score)
```

---

7. Comparing Critic score and User Score to see if they align with each other

```
summary(df_cleaned$Critic_Score)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    13.00   62.00   72.00   70.27   80.00   98.00
```

```
summary(df_cleaned$User_Score)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##     0.000   6.000   7.000   6.737   8.000   9.000
```

---

8. Here we can see that the user score is out of 10 whereas, the Critic score is out of 100. We need to have both the columns in the same range to make useful insights.

```
# So we are converting the critic score range from 0 to 10.
df_cleaned$Critic_Score <- df_cleaned$Critic_Score / 10
summary(df_cleaned$Critic_Score)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##     1.300   6.200   7.200   7.027   8.000   9.800
```

```
summary(df_cleaned$User_Score)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##     0.000   6.000   7.000   6.737   8.000   9.000
```

---

9. We have to check if there are any duplicates present in the cleaned dataset.

```
duplicates <- df_cleaned[duplicated(df_cleaned), ]
# Calculate the count of duplicates
count_of_duplicates <- nrow(duplicates)
count_of_duplicates
```

```
## [1] 0
```

10. There are no duplicate observations in the cleaned dataset.

---

11. Let's check if we can find anything wrong with the column 'Rating'.

```
counts <- table(df_cleaned$Rating)
counts
```

```
##
##   AO      E E10+  K-A      M    RP      T
##     1 2082   930      1 1433      1 2377
```

---

12. We can see that there are few ratings like 'AO', 'K-A', and 'RP' which has only one data. We should remove these observations so as to have better results.

```
df_filtered <- df_cleaned[df_cleaned$Rating != "AO",]
df_filtered <- df_filtered[df_filtered$Rating != "RP",]
df_filtered <- df_filtered[df_filtered$Rating != "K-A",]

counts <- table(df_filtered$Rating)
counts
```

```
##
##      E E10+      M      T
## 2082   930 1433 2377
```

---

13. Now that we have cleaned the ratings column, we are now ready for Visualization.

---

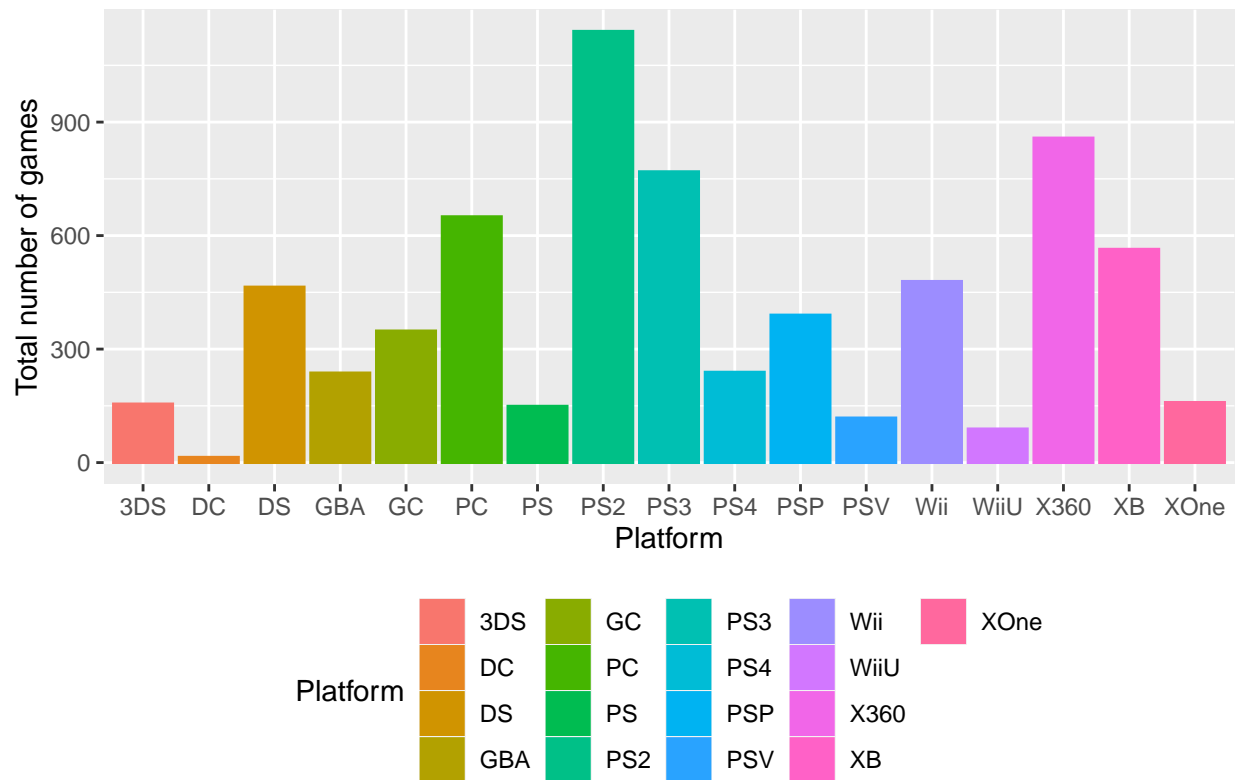
**Problem 2** Visualize something interesting to you from the dataset using ggplot2. Comment on what the visualization shows and any key conclusions.

```
# Importing the library ggplot2 for plotting
library(ggplot2)
```

1. Lets start the visualization from second column, which is "Platform".

```
ggplot(data=df_filtered)+
  geom_bar(aes(x=Platform,color=Platform,fill=Platform)) +
  theme(legend.position="bottom") +
  labs(title = "PS2 has the highest number of video games made",
       x = "Platform",
       y = "Total number of games")
```

## PS2 has the highest number of video games made



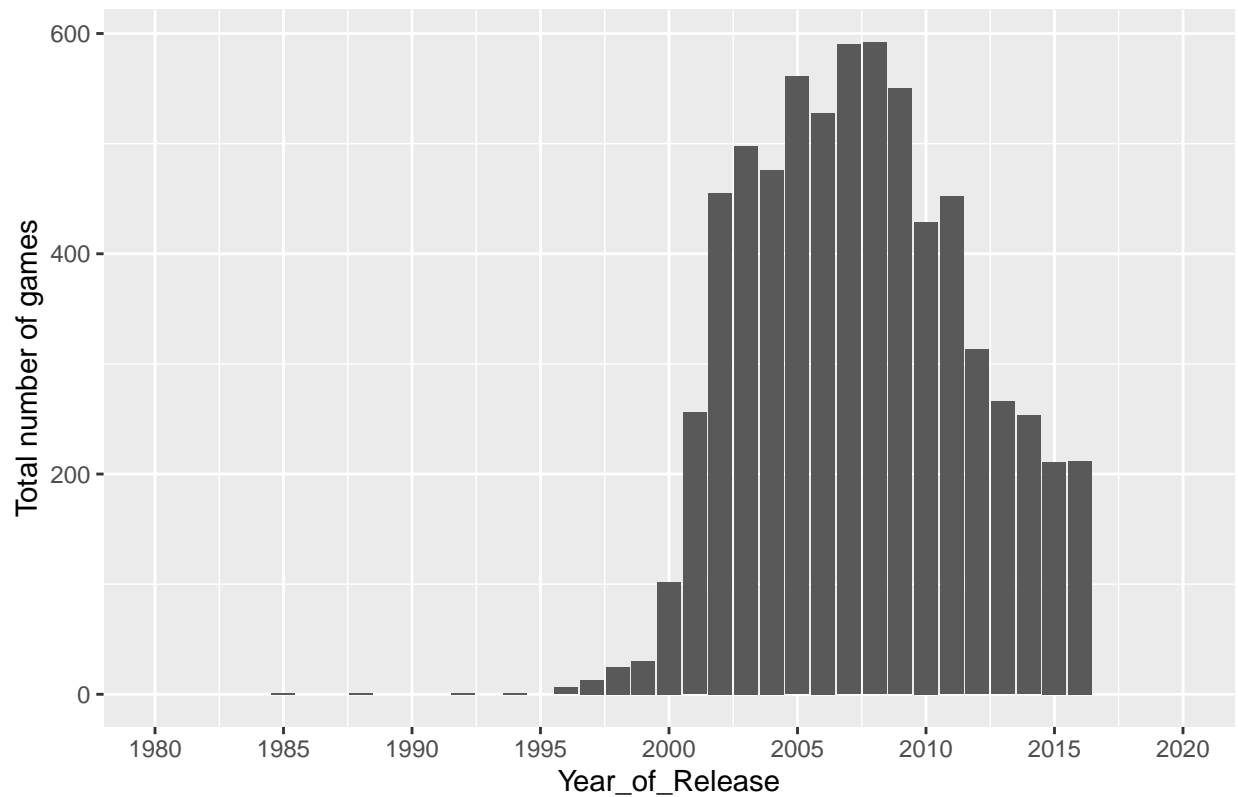
Observation:

- a. PS2 has the highest number of video games made for which user and critics rating are available.

2. Next lets have a bar plot for year of release.

```
ggplot(data=df_filtered)+
  geom_bar(aes(x=Year_of_Release)) +
  theme(legend.position="bottom") +
  scale_x_continuous(limits = c(1980, 2020),
                     breaks = seq(1980, 2020, by = 5)) +
  labs(title = "Year 2008 and 2007 has the highest number of video games released.",
       x = "Year_of_Release",
       y = "Total number of games")
```

Year 2008 and 2007 has the highest number of video games released.



Observations:

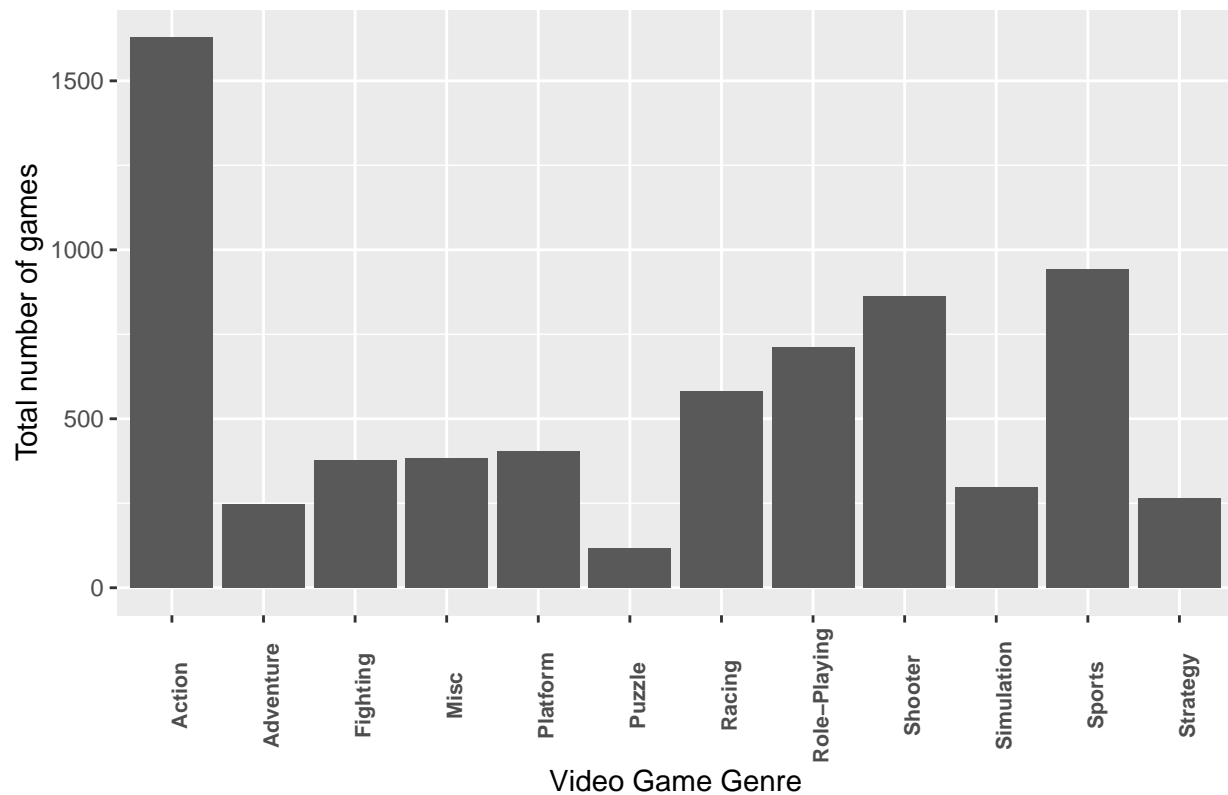
- Year 2007-2008 had the highest number of games released, with a combined total of around 1180 games.
- We can see a sudden jump in the number of games released after 2000, the trend goes negative after 2008.

### 3. Plotting the variable 'Genre'

```
ggplot(data=df_filtered,fig.align="center",fig.width = 50)+
  geom_bar(aes(x=Genre)) +
  theme(legend.position="bottom") +
  theme(axis.text.x = element_text(size = 8, angle = 90, face = "bold")) +
  labs(title = "Action genre is the most famous genre for video games",
       x = "Video Game Genre",
       y = "Total number of games")
```



Action genre is the most famous genre for video games



Observation:

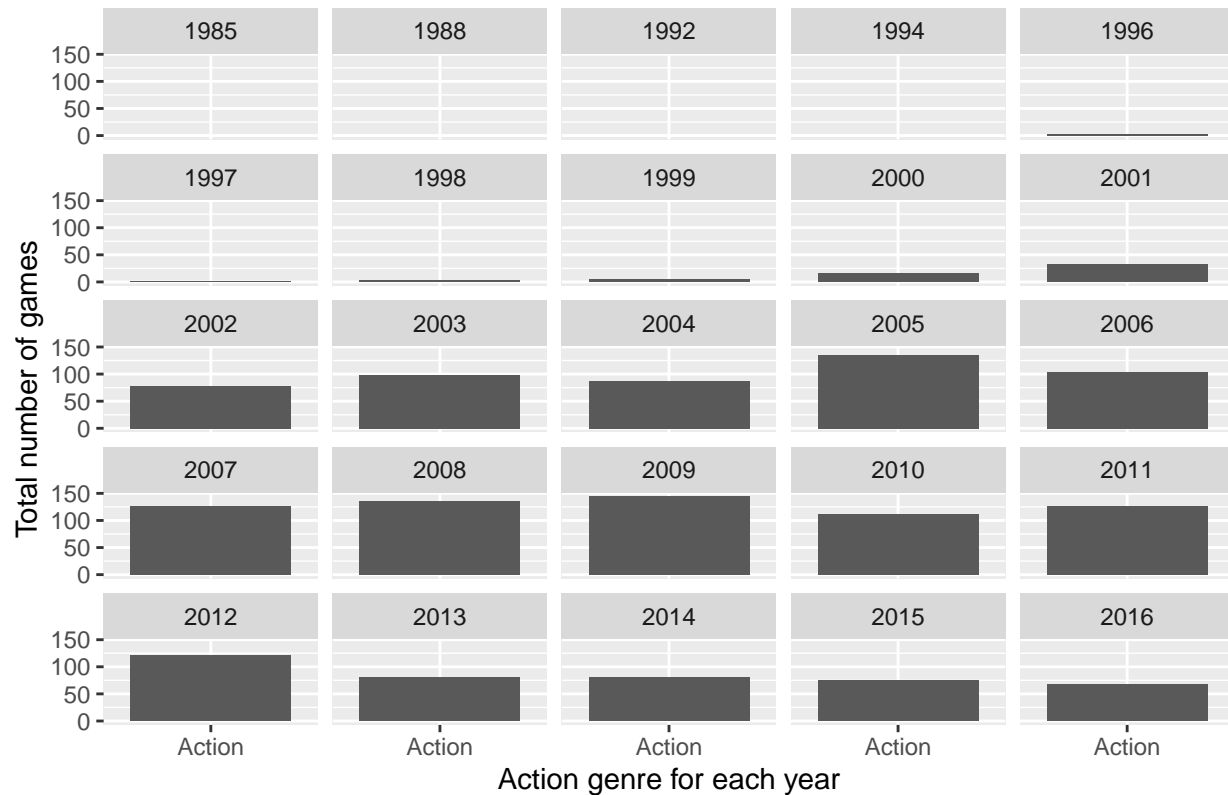
- Genre Action has the highest number of games released, followed by genre Sports.
- Genre Puzzle has the least number of games released.

---

4. Let's try to visualize how many games of the genre 'action' were released in each year.

```
ggplot(data=df_filtered,fig.align="center",fig.width = 50,aes(x=Genre))+  
  geom_bar(data = subset(df_filtered, Genre == 'Action')) + facet_wrap(~Year_of_Release) +  
  theme(legend.position="bottom") +  
  labs(title = "Year 2009 has the highest number of Action games released.",  
        x = "Action genre for each year",  
        y = "Total number of games")
```

Year 2009 has the highest number of Action games released.



Observation:

1. The year 2009 has the most number of action games released.

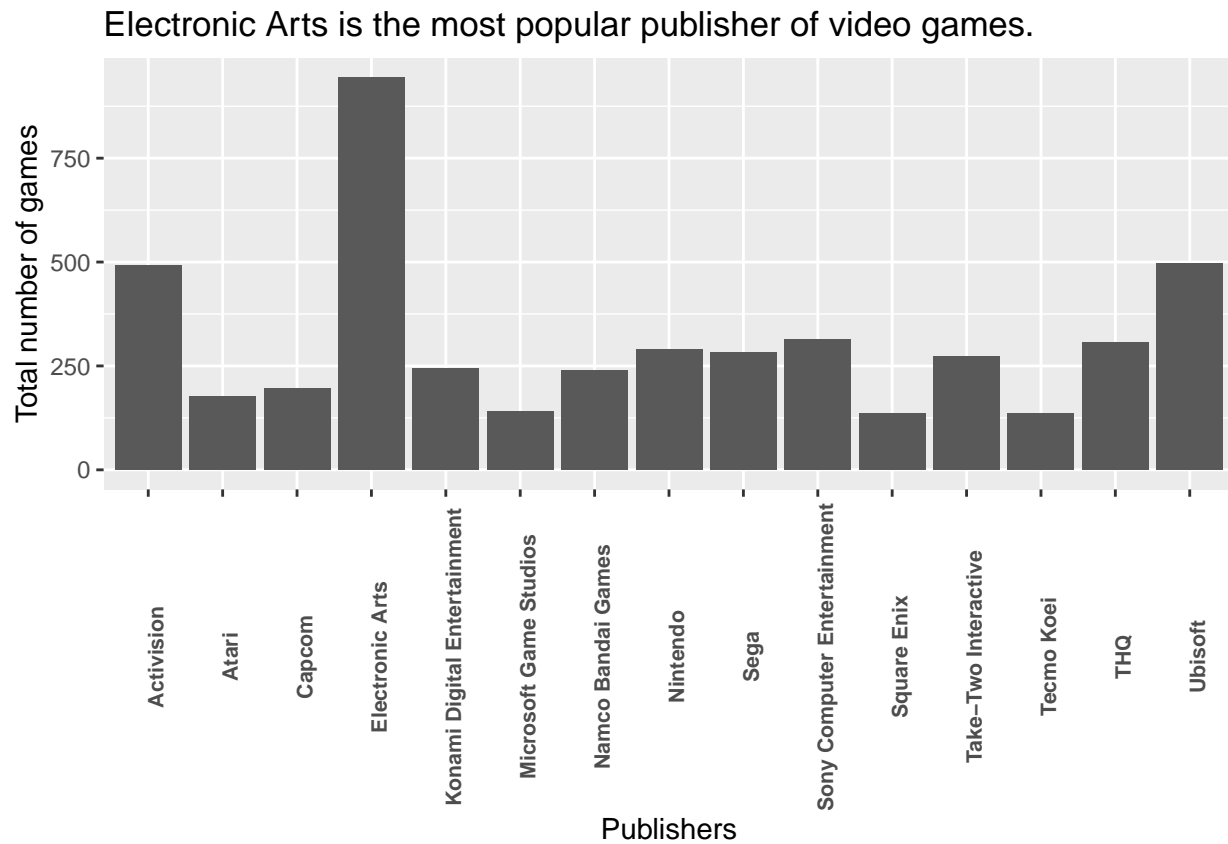
4. Let's see what we can find by visualizing the top 15 publishers with the highest games published.

```
xy <- sort(table(df_filtered$Publisher),decreasing = TRUE)
publisher_counts <- sort(table(df_filtered$Publisher), decreasing = TRUE)
top_publishers <- names(publisher_counts)[1:15]
df_filtered2 <- df_filtered[df_filtered$Publisher %in% top_publishers, ]
top_publishers
```

```
## [1] "Electronic Arts" "Ubisoft"
## [3] "Activision" "Sony Computer Entertainment"
## [5] "THQ" "Nintendo"
## [7] "Sega" "Take-Two Interactive"
## [9] "Konami Digital Entertainment" "Namco Bandai Games"
## [11] "Capcom" "Atari"
## [13] "Microsoft Game Studios" "Tecmo Koei"
## [15] "Square Enix"
```

Plotting a bar graph to visualize the data for top 15 publishers.

```
ggplot(data=df_filtered2,aes(x=Publisher))+
  geom_bar() +
  theme(legend.position="bottom") +
  theme(axis.text.x = element_text(size = 8, angle = 90, face = "bold")) +
  labs(title = "Electronic Arts is the most popular publisher of video games.",
       x = "Publishers",
       y = "Total number of games")
```



Observations:

- a. Electronic Arts is the most popular publisher of video games, with more than 825 games published.

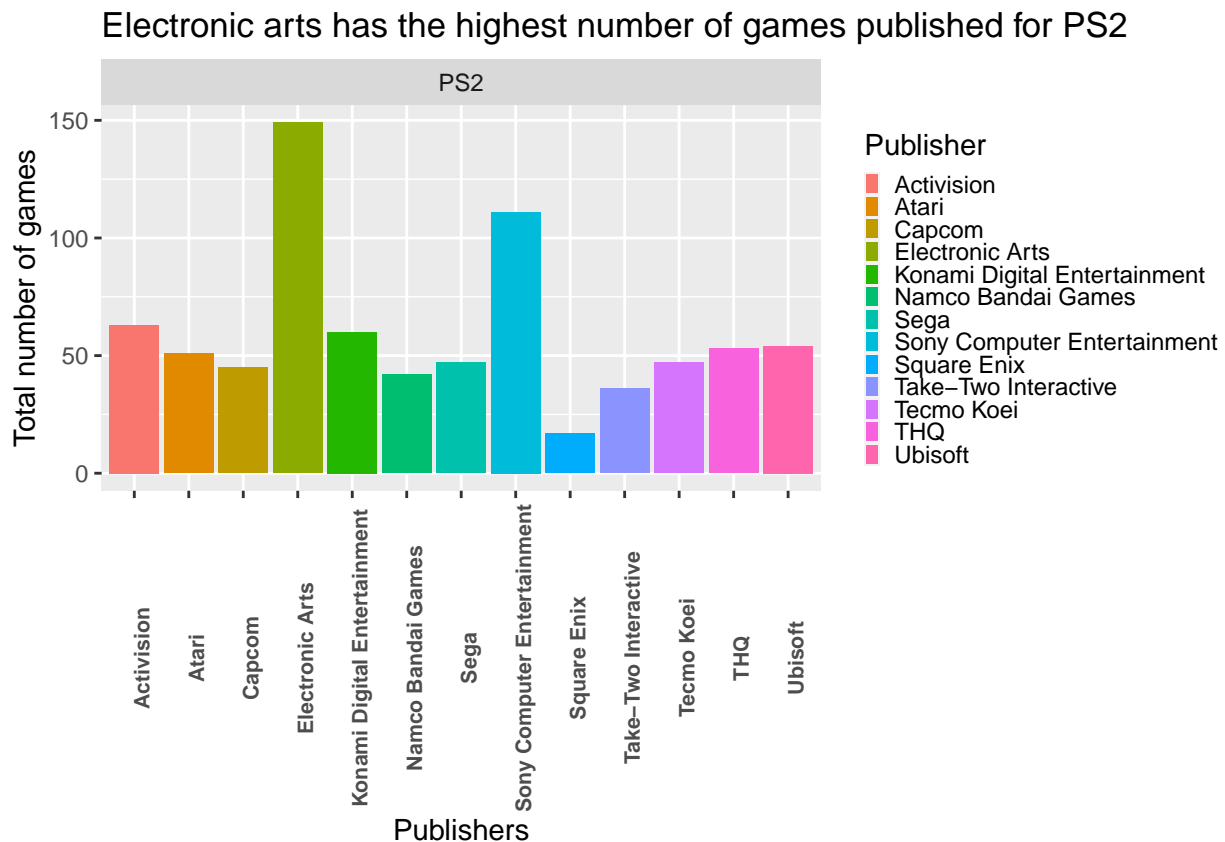
5. Lets see if we can find any relationship between the most common gaming console that is ps2 and top 15 publishers. For this, we have to then update the df\_filtered2 to have only the PS2 data.

```
df_ps2 <- df_filtered2[df_filtered2$Platform == "PS2",]
```

So now we have all the data with platform as PS2 and for the top 15 publishers.

```
ggplot(data=df_ps2,fig.align="center",fig.width = 50,aes(x=Publisher))+
  geom_bar(aes(fill=Publisher)) + facet_wrap(~Platform) +
  theme(axis.text.x = element_text(size = 8, angle = 90, face = "bold")) +
```

```
labs(title = "Electronic arts has the highest number of games published for PS2",
     x = "Publishers",
     y = "Total number of games") + theme(legend.key.size = unit(0.2, "cm"))
```



Observations:

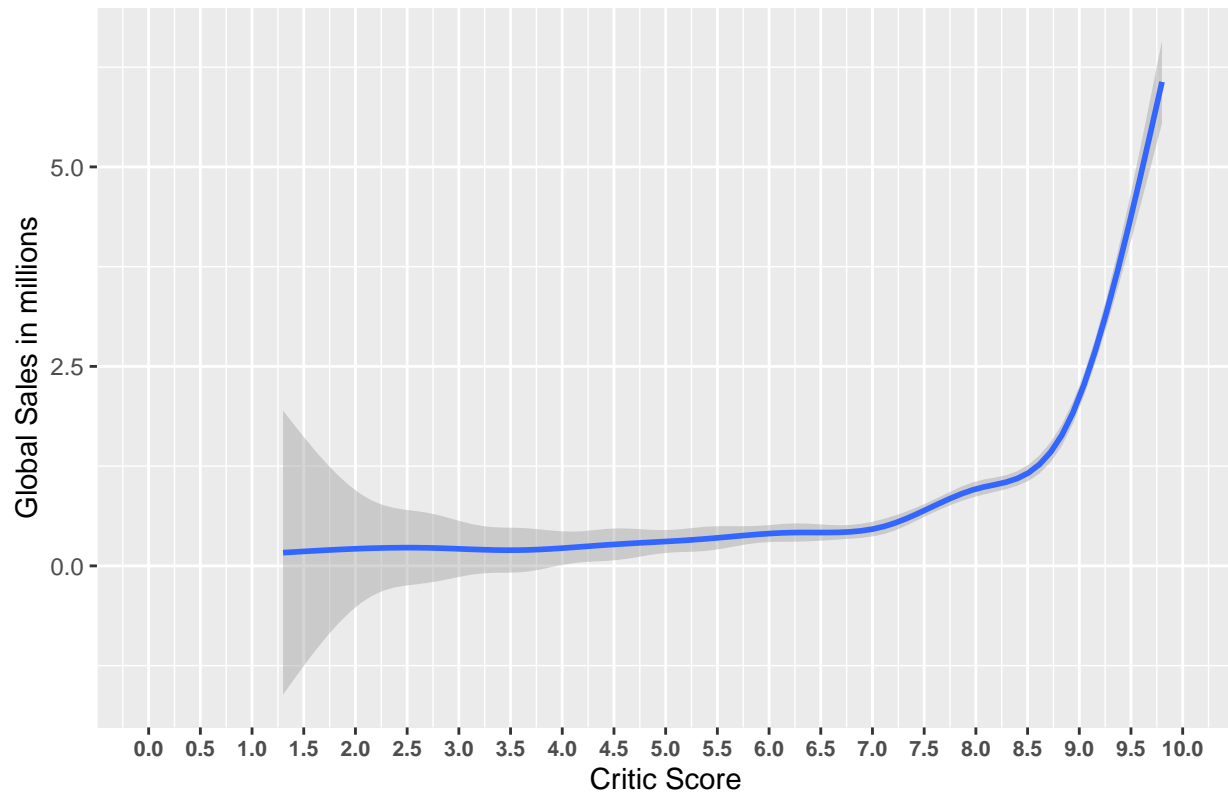
1. Electronic arts is still the most famous video game publisher for the platform 'PS2'.
2. Sony Computer Entertainment stands second which was expected as the platform 'PS2' is developed by Sony.

---

6. Next we have sales columns, lets plot each sales column in a scatter plot

```
ggplot(data=df_filtered,aes(x=Critic_Score,y=Global_Sales))+ geom_smooth() +
  theme(legend.position="bottom") +
  scale_x_continuous(limits = c(0,10),
                    breaks = seq(0,10, by = 0.5)) +
  theme(axis.text.x = element_text(size = 8, face = "bold")) +
  labs(title = "Global Sales increases sharply after the critic score of 8.5",
       x = "Critic Score",
       y = "Global Sales in millions")
```

Global Sales increases sharply after the critic score of 8.5



Observations:

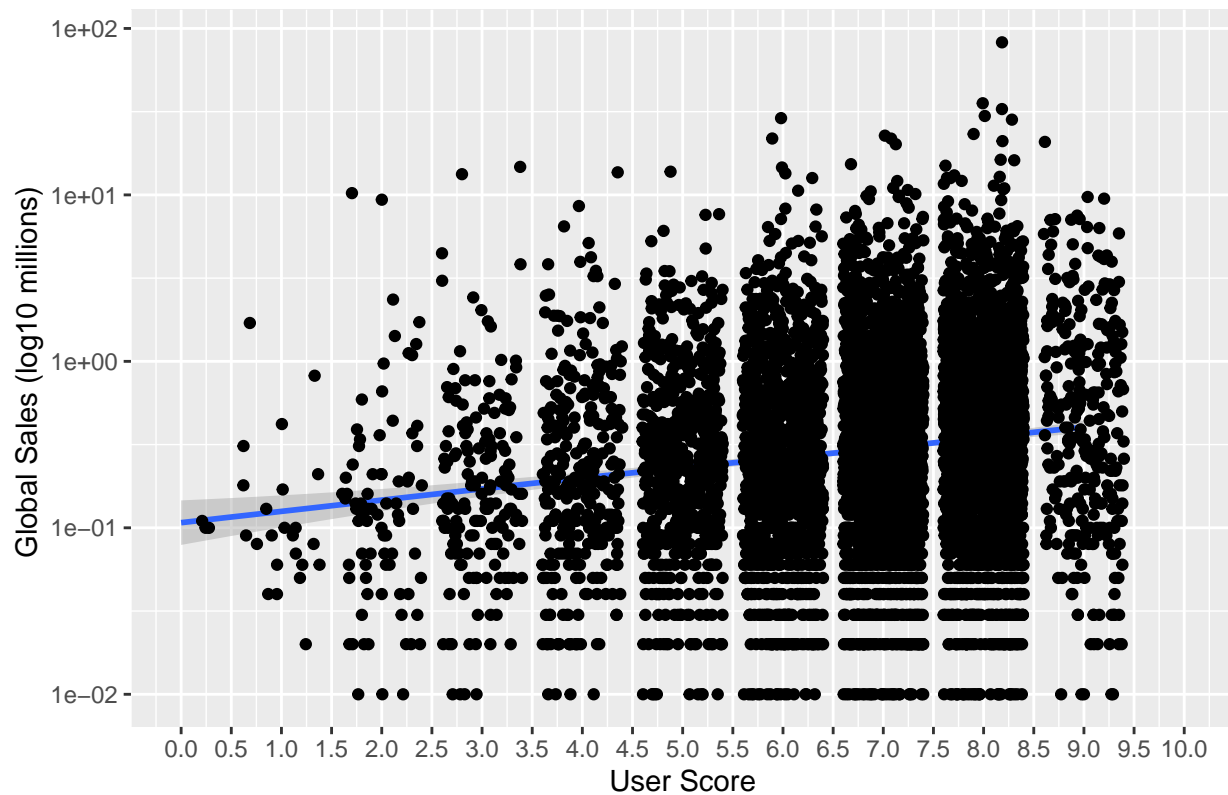
- Global Sales and Critic Score has a positive relationship, with sales increasing as the critic score increases.
- There is a sharp rise in the global sales of the video games after the critic score of 8.5.
- The least critic score is somewhere between 1.25 - 1.3 and the highest critic score is in the range 9.75 - 9.8 .

---

7. Plotting user score against global sales

```
ggplot(data=df_filtered,aes(x=User_Score,y= Global_Sales))+  
  geom_smooth() +geom_jitter() + scale_y_log10() +  
  scale_x_continuous(limits = c(0,10),  
                     breaks = seq(0,10, by = 0.5)) +  
  labs(title = "Global Sales increases as the User Score increases",  
        x = "User Score",  
        y = "Global Sales (log10 millions) ")
```

## Global Sales increases as the User Score increases



Observation:

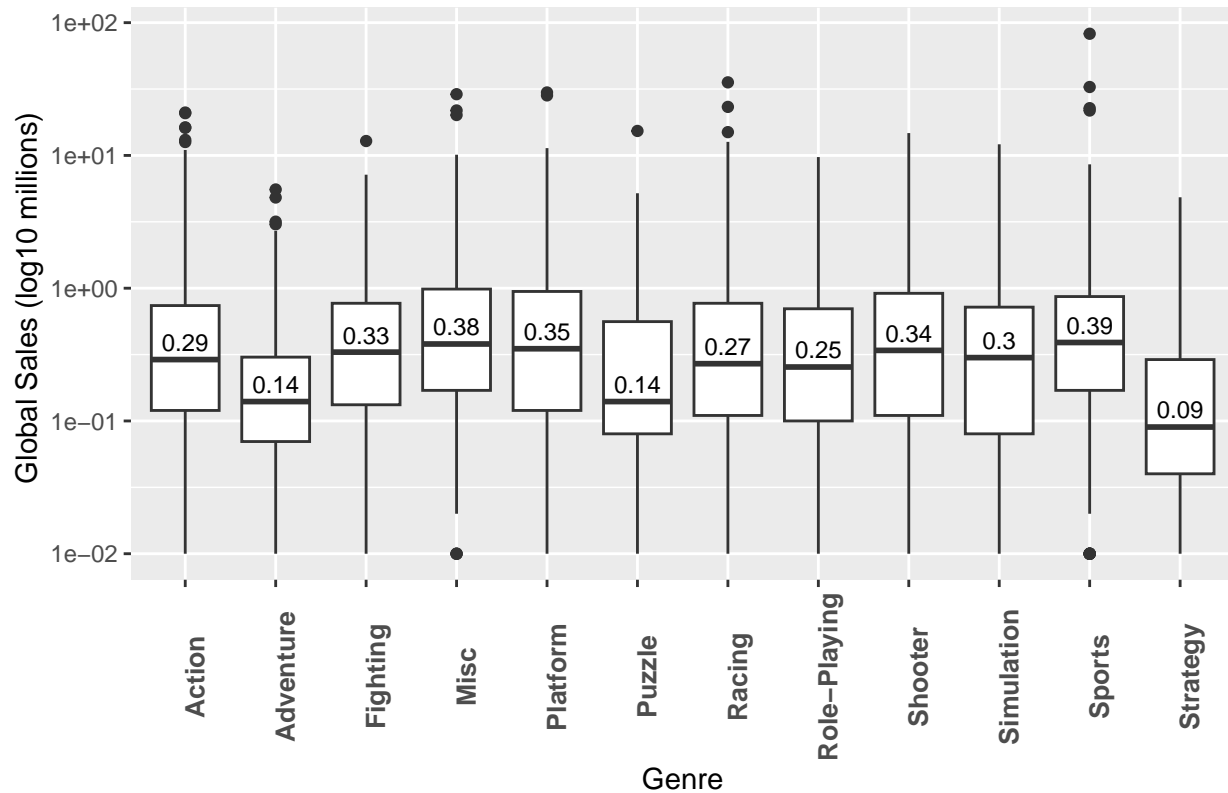
- Here also we see a general trend of global sales increasing as the user score increases.
- There are very few games with a critic score of more than 8.5, whereas there are no games with user score more than 9.5. This tells that users are more strict in giving scores in compare to the critics.
- The range 6.5 - 8.5, is the most dense.

---

8. Let's plot global sales along with genre of the game to see which genre is actually sold the highest.

```
df_filtered %>%
  ggplot(aes(x=Genre, y= Global_Sales)) +
  geom_boxplot() + scale_y_log10() +
  stat_summary(fun = median, geom = "text",
               aes(label = round(..y.., 2)),
               vjust = -0.5, size = 3)+
  theme(plot.title = element_text(hjust=0.5)) +
  theme(axis.text.x = element_text(size = 10, angle = 90, face = "bold")) +
  labs(title = "Genre Sports has the highest number of global sales.",
       x = "Genre",
       y = "Global Sales (log10 millions) ")
```

Genre Sports has the highest number of global sales.



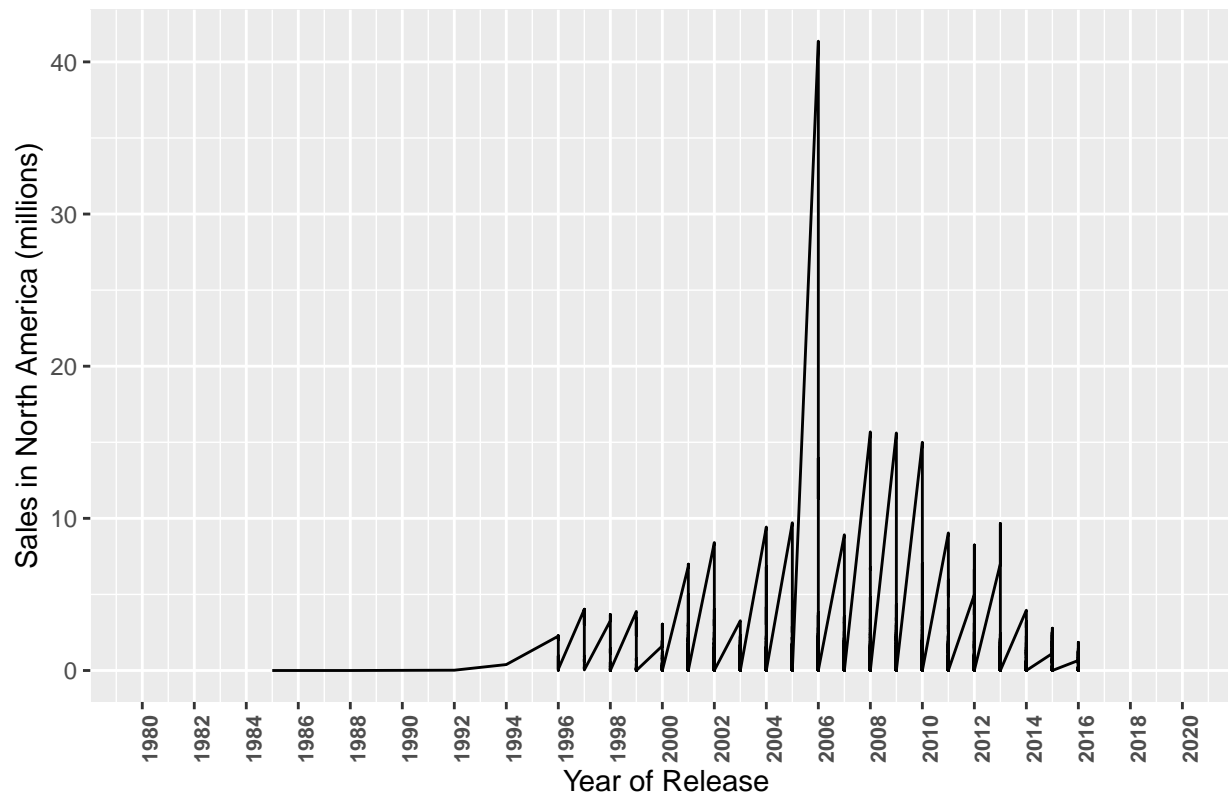
Observation:

- Here we can see that genre like puzzle, strategy, and adventure have really low count of sales in compare to other genre.
- Sports genre has the highest number of sales, with the median value around 0.39 in log10 millions.

9. Plotting sales from each continent to see how they varied from time

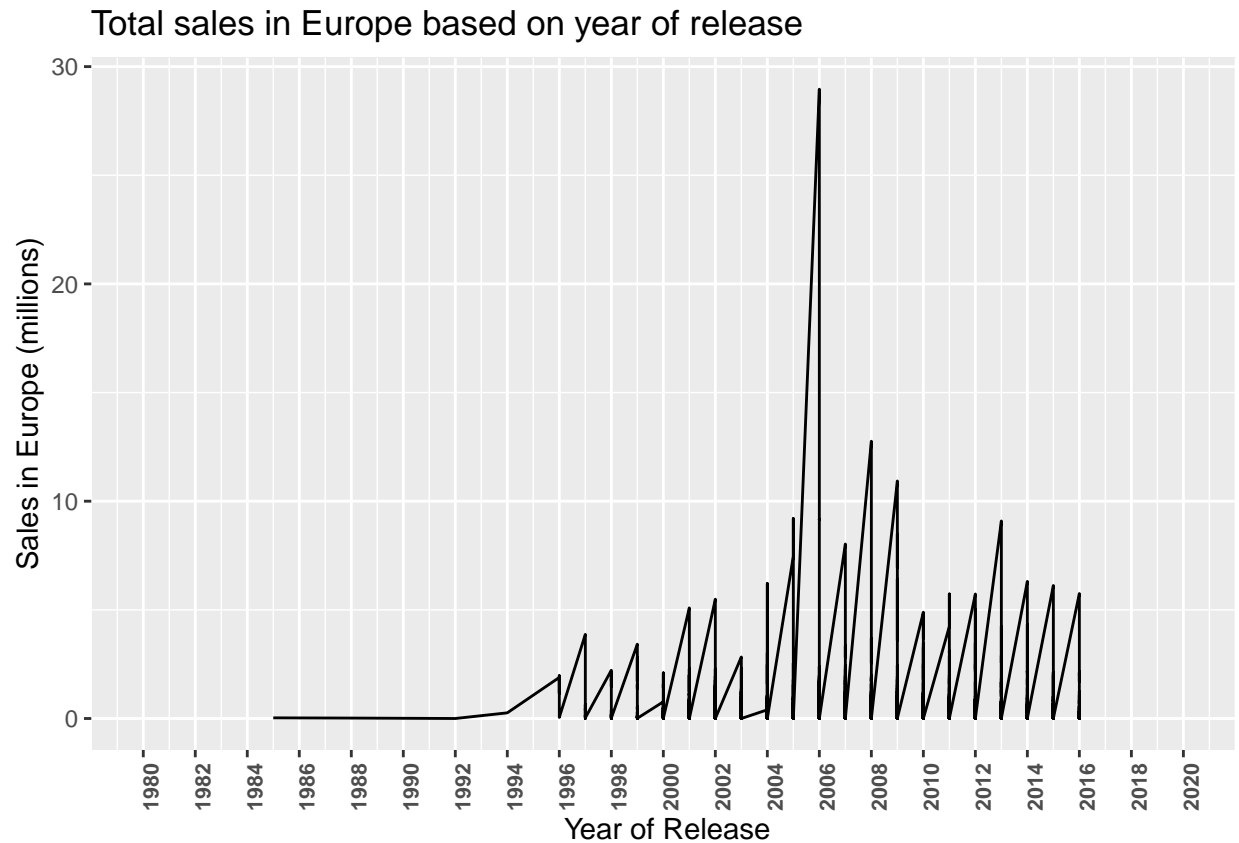
```
ggplot(data=df_filtered,fig.align="center",fig.width = 50)+
  geom_line(aes(x=Year_of_Release, y= NA_Sales)) +
  scale_x_continuous(limits = c(1980, 2020),
                     breaks = seq(1980, 2020, by = 2)) +
  theme(legend.position="bottom") +
  theme(axis.text.x = element_text(size = 8, angle = 90, face = "bold")) +
  labs(title = "Total sales in North America based on year of release",
       x = " Year of Release",
       y = "Sales in North America (millions)")
```

Total sales in North America based on year of release



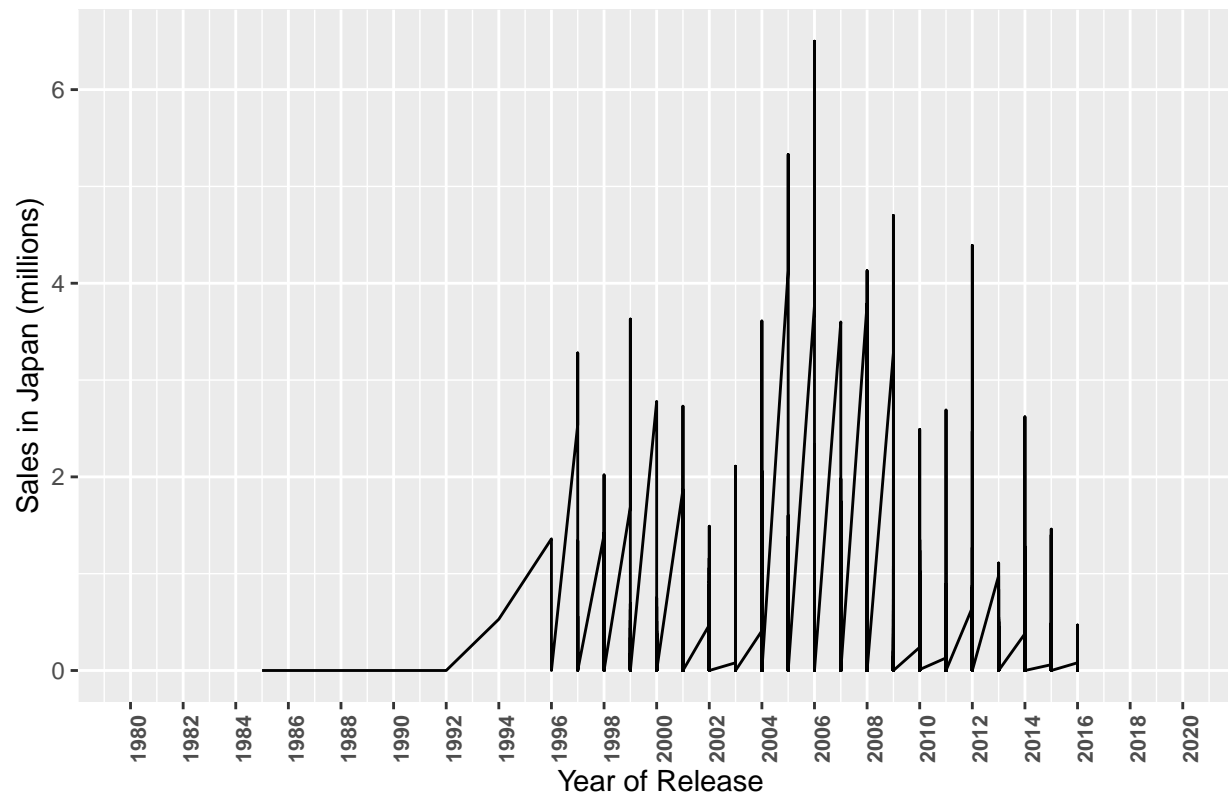
```
ggplot(data=df_filtered,fig.align="center",fig.width = 50)+
  geom_line(aes(x=Year_of_Release, y= EU_Sales))  +
  theme(legend.position="bottom") +
  scale_x_continuous(limits = c(1980, 2020),
                     breaks = seq(1980, 2020, by = 2)) +
  theme(axis.text.x = element_text(size = 8, angle = 90, face = "bold"))+
  labs(title = "Total sales in Europe based on year of release",
       x = " Year of Release",
       y = "Sales in Europe (millions)")
```





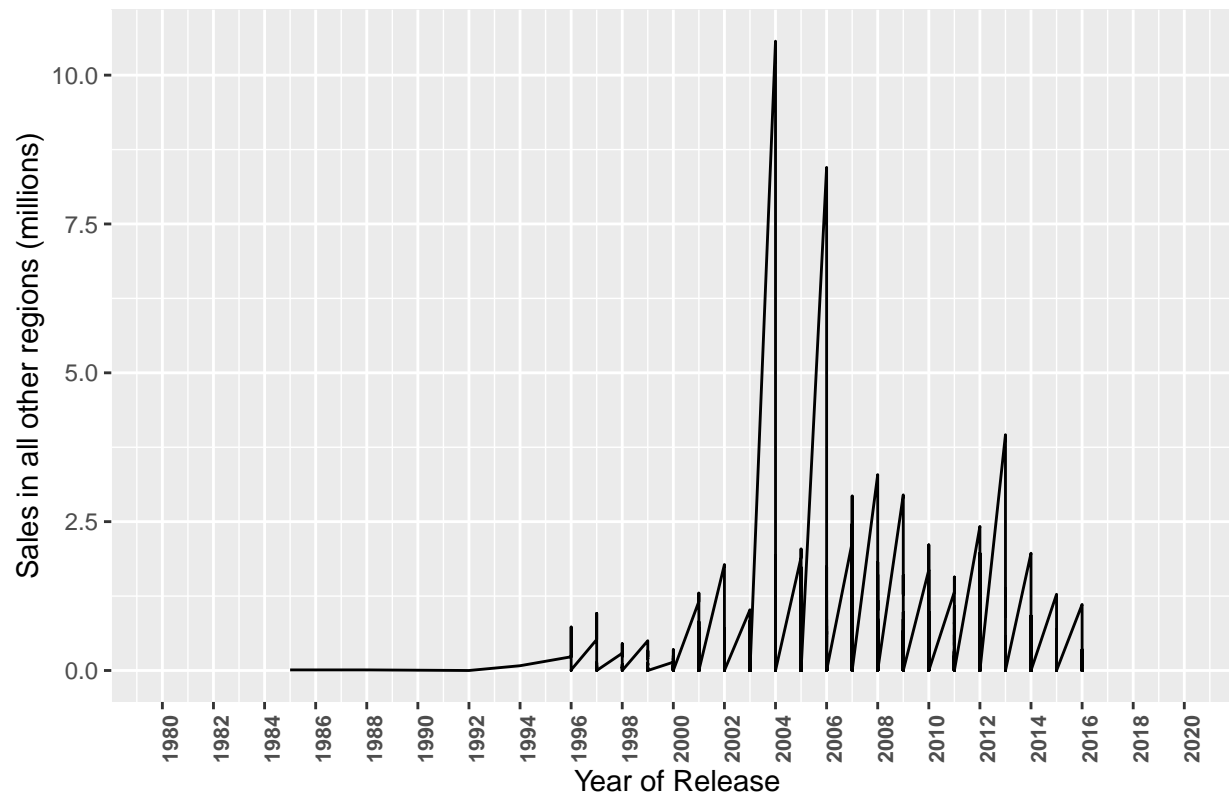
```
ggplot(data=df_filtered,fig.align="center",fig.width = 50)+
  geom_line(aes(x=Year_of_Release, y= JP_Sales))  +
  theme(legend.position="bottom") +
  scale_x_continuous(limits = c(1980, 2020),
                     breaks = seq(1980, 2020, by = 2)) +
  theme(axis.text.x = element_text(size = 8, angle = 90, face = "bold")) +
  labs(title = "Total sales in Japan based on year of release",
       x = " Year of Release",
       y = "Sales in Japan (millions)")
```

Total sales in Japan based on year of release



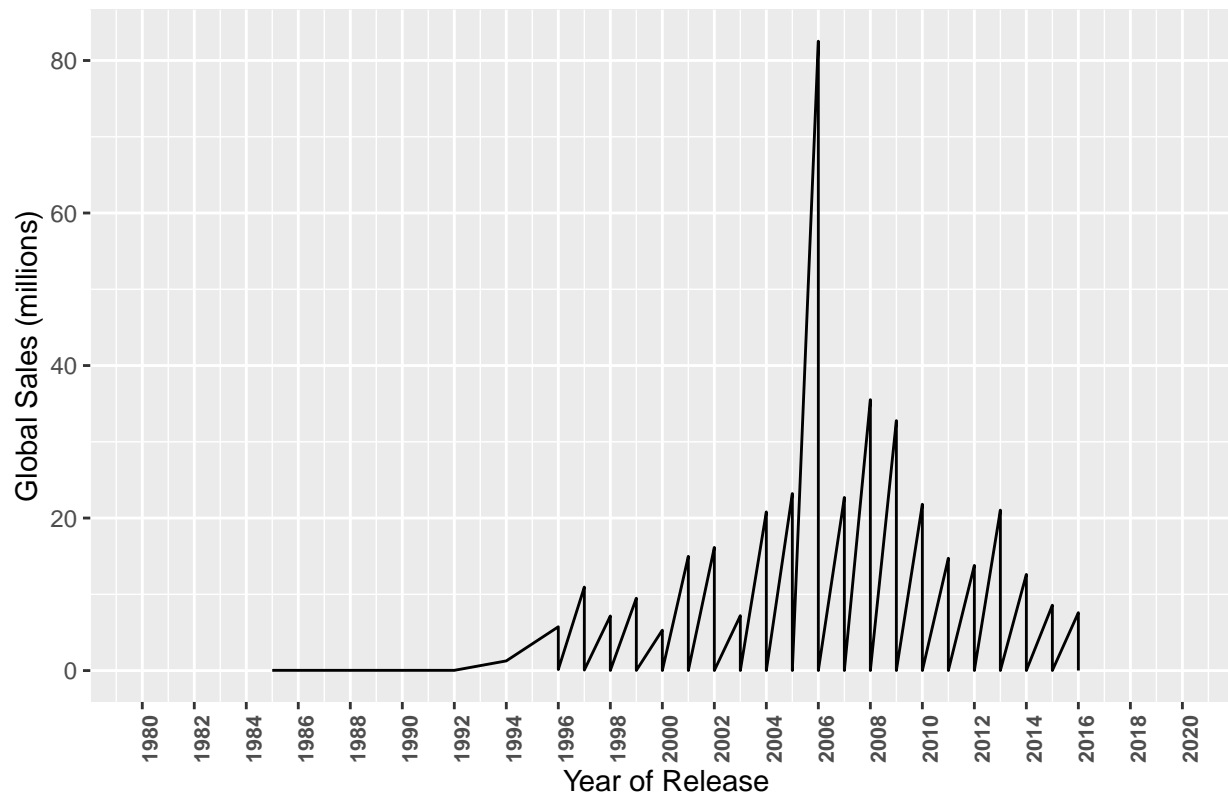
```
ggplot(data=df_filtered,fig.align="center",fig.width = 50)+
  geom_line(aes(x=Year_of_Release, y= Other_Sales)) +
  theme(legend.position="bottom") +
  scale_x_continuous(limits = c(1980, 2020),
                     breaks = seq(1980, 2020, by = 2)) +
  theme(axis.text.x = element_text(size = 8, angle = 90, face = "bold")) +
  labs(title = "Total sales in other regions based on year of release",
       x = " Year of Release",
       y = "Sales in all other regions (millions)")
```

Total sales in other regions based on year of release



```
ggplot(data=df_filtered,fig.align="center",fig.width = 50)+
  geom_line(aes(x=Year_of_Release, y= Global_Sales))  +
  theme(legend.position="bottom") +
  scale_x_continuous(limits = c(1980, 2020),
                     breaks = seq(1980, 2020, by = 2)) +
  theme(axis.text.x = element_text(size = 8, angle = 90, face = "bold")) +
  labs(title = "Global sales based on year of release",
       x = " Year of Release",
       y = "Global Sales (millions)")
```

Global sales based on year of release



Observation:

- a. Sales in North America, Europe and globally combined increased rapidly in the year 2005, which followed a sharp fall in the same year as well.
- b. In Japan, there was a continue increase and decrease in the video games sales, with the highest games sold in the year 2005 - 2006.
- c. Total Sales in other regions were highest in the year 2003 - 2004.

---

10. To research more on what genre of games are sold between the year 1995-2015 in Japan, we'll have facet of top genre played in japan.

```
# we want to see how top 5 genres of global sales did in japan in the year 1996-2015
# so first have a subset of data of the games sold between the year 1995-2015
```

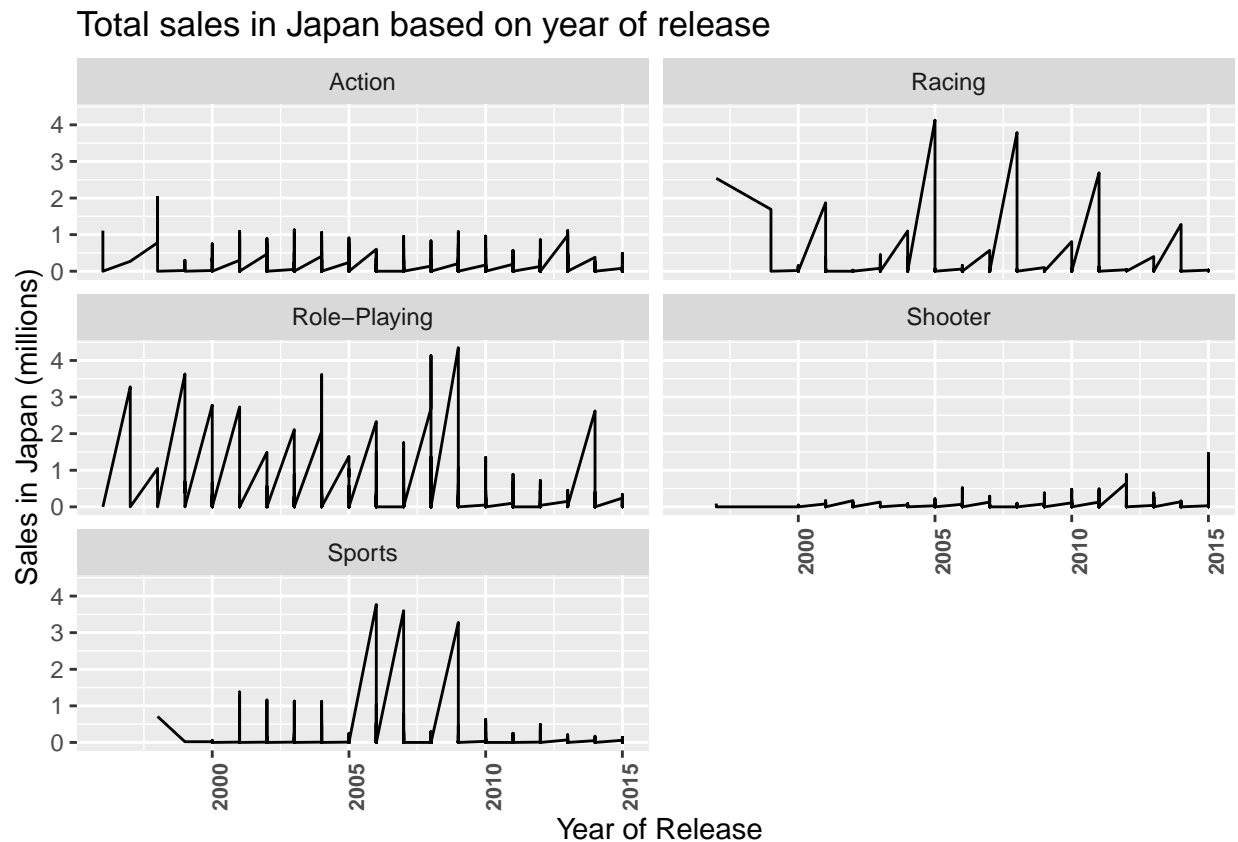
```
df_tp <- df_filtered[df_filtered$Year_of_Release > 1995,]
df_tp2 <- df_tp[df_tp$Year_of_Release < 2016,]

# df_tp2 consists of data between the year 1995-2015
xy <- sort(table(df_tp2$Genre), decreasing = TRUE)
genre_counts <- sort(table(df_tp2$Genre), decreasing = TRUE)
top_genre <- names(genre_counts)[1:5]
```

```
df_tp3 <- df_tp2[df_tp2$Genre %in% top_genre, ]

# Now that we have top 5 genres played in the year 1995-2015, we can plot a graph to
# see how each of these genre did in Japan

ggplot(data=df_tp3,fig.align="center",fig.width = 50)+
  geom_line(aes(x=Year_of_Release, y= JP_Sales)) + facet_wrap(~Genre,nrow=3) +
  theme(legend.position="bottom") +
  theme(axis.text.x = element_text(size = 8, angle = 90, face = "bold")) +
  labs(title = "Total sales in Japan based on year of release",
       x = "Year of Release",
       y = "Sales in Japan (millions)")
```



Observations:

- a. Genre Role-Playing has the highest number of sales in Japan, with the highest number of games sold in the year 2008-2009.
- b. Genre Shooting has the lowest number of sales in Japan, with the highest sold in the year 2015.

---

11. Let's plot the developer column to see the top 15 developers.

```

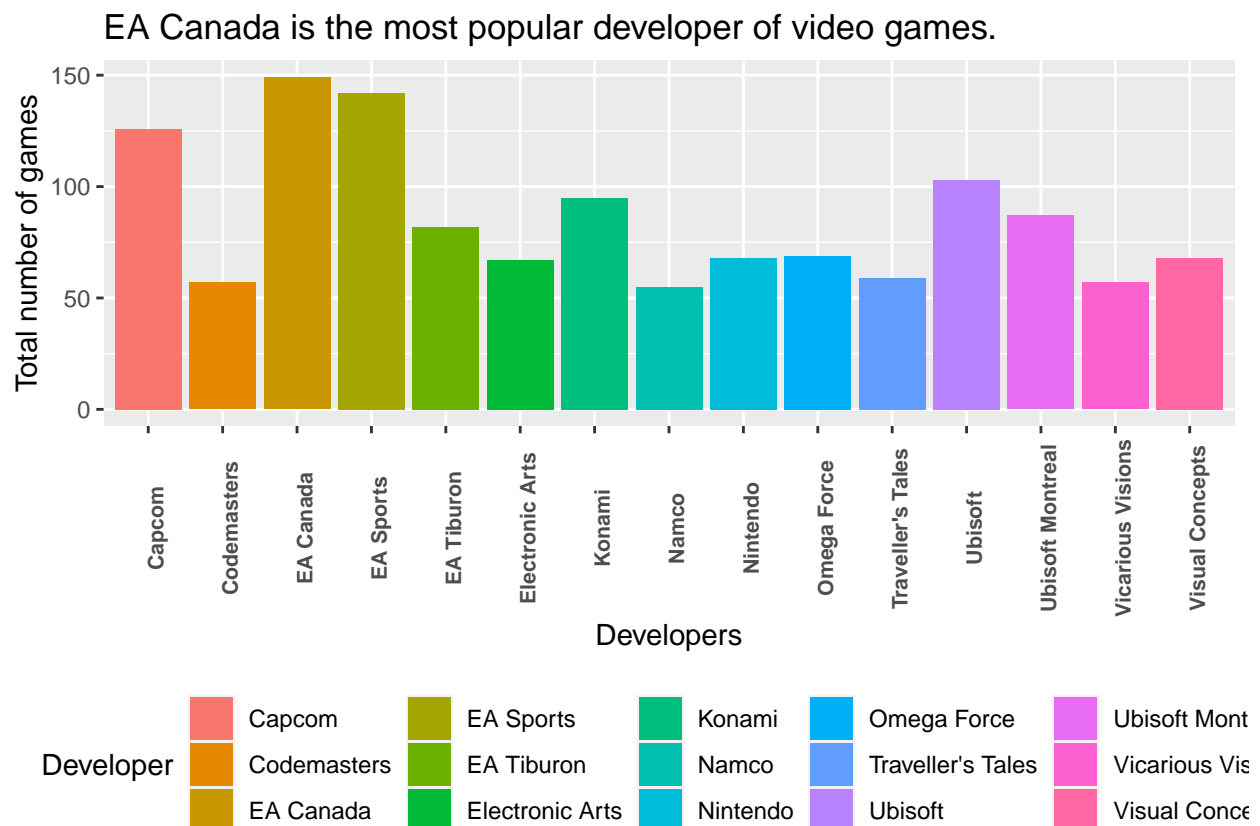
xy <- sort(table(df_filtered$Developer),decreasing = TRUE)

developer_counts <- sort(table(df_filtered$Developer), decreasing = TRUE)
top_developers <- names(developer_counts)[1:15]

df_filtered3 <- df_filtered[df_filtered$Developer %in% top_developers, ]

ggplot(data=df_filtered3,aes(x=Developer))+
  geom_bar(aes(fill=Developer)) +
  theme(legend.position="bottom") +
  theme(axis.text.x = element_text(size = 8, angle = 90, face = "bold")) +
  labs(title = "EA Canada is the most popular developer of video games.",
       x = "Developers",
       y = "Total number of games")

```



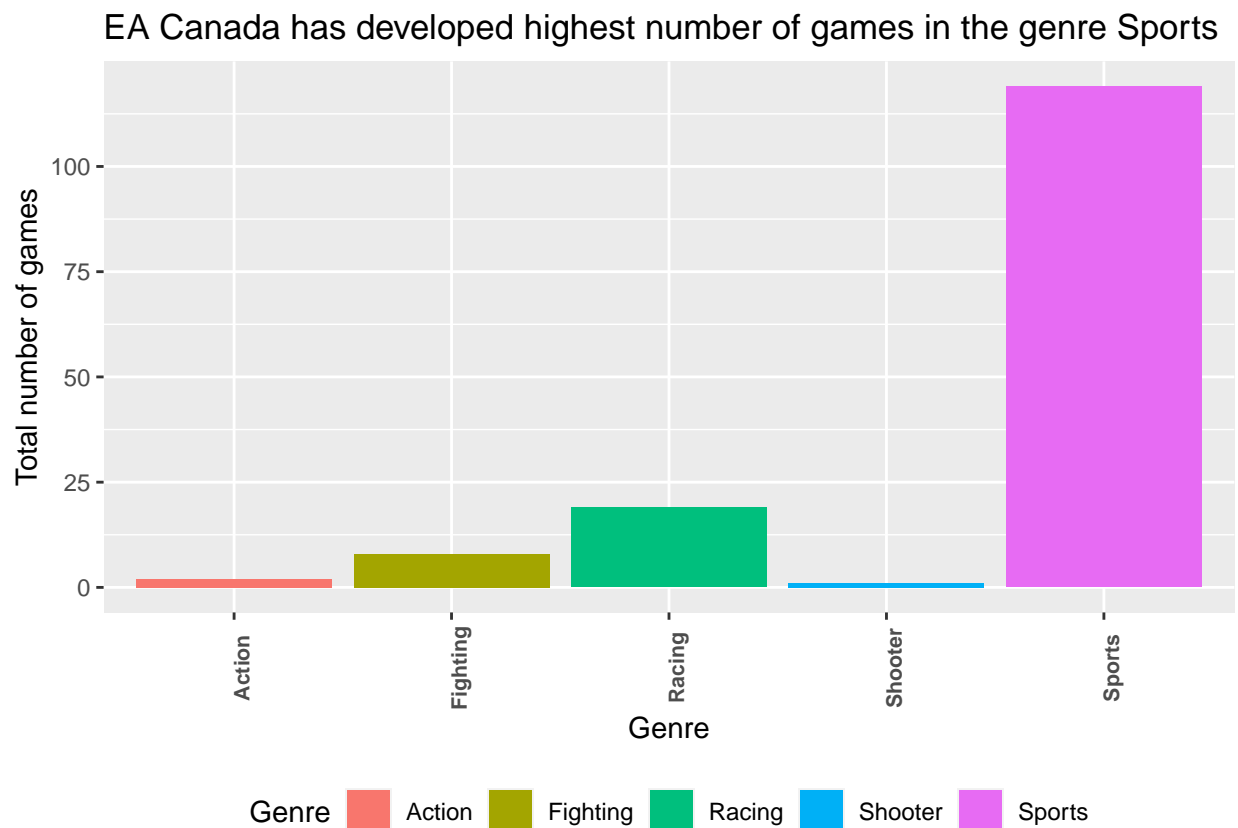
Observations:

- EA Canada tops the chart following by EA Sports.

12. Let's try to visualize the what are the different types of genres on which video games have been developed by EA Canada.

```
dfEACanada <- df_filtered[df_filtered$Developer == "EA Canada",]
```

```
ggplot(data=dfEACanada,aes(x=Genre))+
  geom_bar(aes(fill=Genre)) +
  theme(legend.position="bottom") +
  theme(axis.text.x = element_text(size = 8, angle = 90, face = "bold")) +
  labs(title = "EA Canada has developed highest number of games in the genre Sports",
       x = "Genre",
       y = "Total number of games")
```



Observation:

- a. EA Canada has developed the highest number of games in the genre Sports, and the least in genre Shooter.

---

13. Let Visualize how many games are developed for each rating.

```
plot <- ggplot(data=df_filtered,aes(x=Rating))+
  geom_bar(aes(fill=Rating)) +
  theme(legend.position="bottom") +
  theme(axis.text.x = element_text(size = 10, face = "bold")) +
```

```
labs(title = "ESRB rating 'T' has the highest number of games developed.",
      x = "ESRB Rating",
      y = "Total number of games") +
png("ESRBrating.png", width = 800, height = 600, res = 120)
print(plot)
dev.off()
```

```
## pdf
## 2
```

Observations:

- a. ESRB rating 'T' has the highest number of games developed.
- b. ESRB rating 'E10+' has the lowest number of games developed.

---

13. Let's visualize which rating has the highest number of global sales.

```
plot<- ggplot(data=df_tp3,fig.align="center",fig.width = 50)+
  geom_line(aes(x=Year_of_Release, y= Global_Sales)) + facet_wrap(~Rating) +
  theme(legend.position="bottom") +
  theme(axis.text.x = element_text(size = 8, angle = 90, face = "bold")) +
  labs(title = "ESRB Rating 'E' has the highest global sales.",
        x = "Year of Release",
        y = "Global Sales (millions)")
png("ESRBratingSales.png", width = 800, height = 600, res = 120)
print(plot)
dev.off()
```

```
## pdf
## 2
```

Observations:

- a. ESRB rating 'E' has the highest global sales.
- b. ESRB rating 'T' is the rating with the highest number of games developed, but they have very less sale.