Video Game Analysis

"Which game to play?" As a genuine game lover, this is my topic for the analysis. By studying the sales factors of three regions (NA, EU, JP) discover the most popular game genres around the world. Review scores of critics and users prove the completenss of the games, and the amount of games published by each publisher shows their reputation.

Sub-Topics

1. Most Popular Game Genres by Region
2. Critic Score and User Score of Genres
3. Publishers with Good Reputation
4. Conclusion

Load File

```
library(ggplot2)
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
```

```
games <- read.csv("Video_Games_Sales_as_at_22_Dec_2016.csv")
head(games, 3)
```

```
##                Name Platform Year_of_Release    Genre Publisher NA_Sales
## 1        Wii Sports      Wii            2006   Sports  Nintendo    41.36
## 2 Super Mario Bros.      NES            1985 Platform  Nintendo    29.08
## 3    Mario Kart Wii      Wii            2008   Racing  Nintendo    15.68
##   EU_Sales JP_Sales Other_Sales Global_Sales Critic_Score Critic_Count
## 1    28.96     3.77        8.45        82.53           76           51
## 2     3.58     6.81        0.77        40.24           NA           NA
## 3    12.76     3.79        3.29        35.52           82           73
##   User_Score User_Count Developer Rating
## 1          8        322  Nintendo      E
## 2                    NA
## 3        8.3        709  Nintendo      E
```

1. Most Popular Game Genres by Region

1-1. Most Popular Game Genres in North America

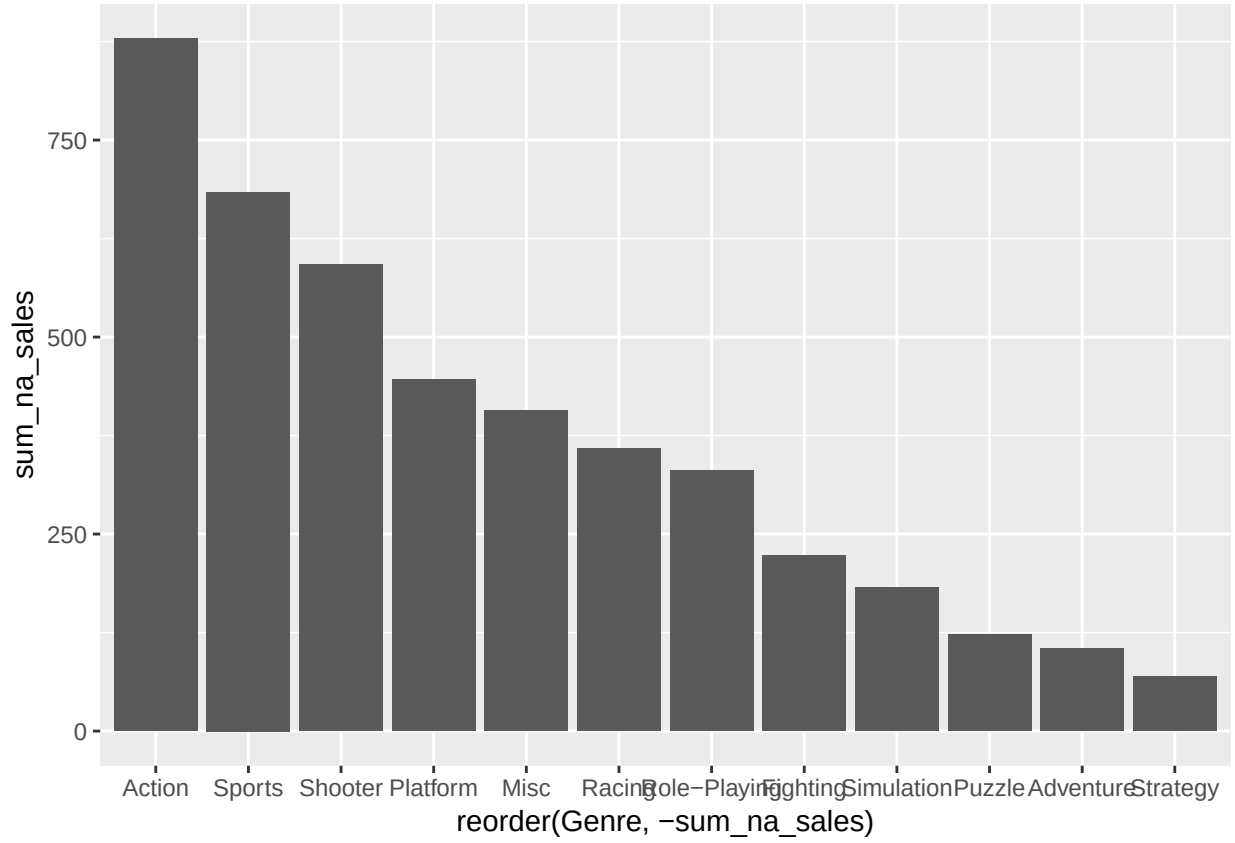
```
genre_na <- games %>%
  group_by(Genre) %>%
  summarise(sum_na_sales = round(sum(NA_Sales), 0)) %>%
```

```
  arrange(desc(sum_na_sales)) %>%
  mutate(na_rank = dense_rank(desc(sum_na_sales))) %>%
  head(12)

genre_na
```

```
## # A tibble: 12 x 3
##    Genre        sum_na_sales na_rank
##    <fct>               <dbl>   <int>
##  1 Action                879       1
##  2 Sports                684       2
##  3 Shooter               592       3
##  4 Platform              446       4
##  5 Misc                  407       5
##  6 Racing                359       6
##  7 Role-Playing          331       7
##  8 Fighting              223       8
##  9 Simulation            182       9
## 10 Puzzle                123      10
## 11 Adventure             105      11
## 12 Strategy               69      12
```

```
ggplot(data=genre_na, aes(x=reorder(Genre, -sum_na_sales), y=sum_na_sales)) +
  geom_col()
```



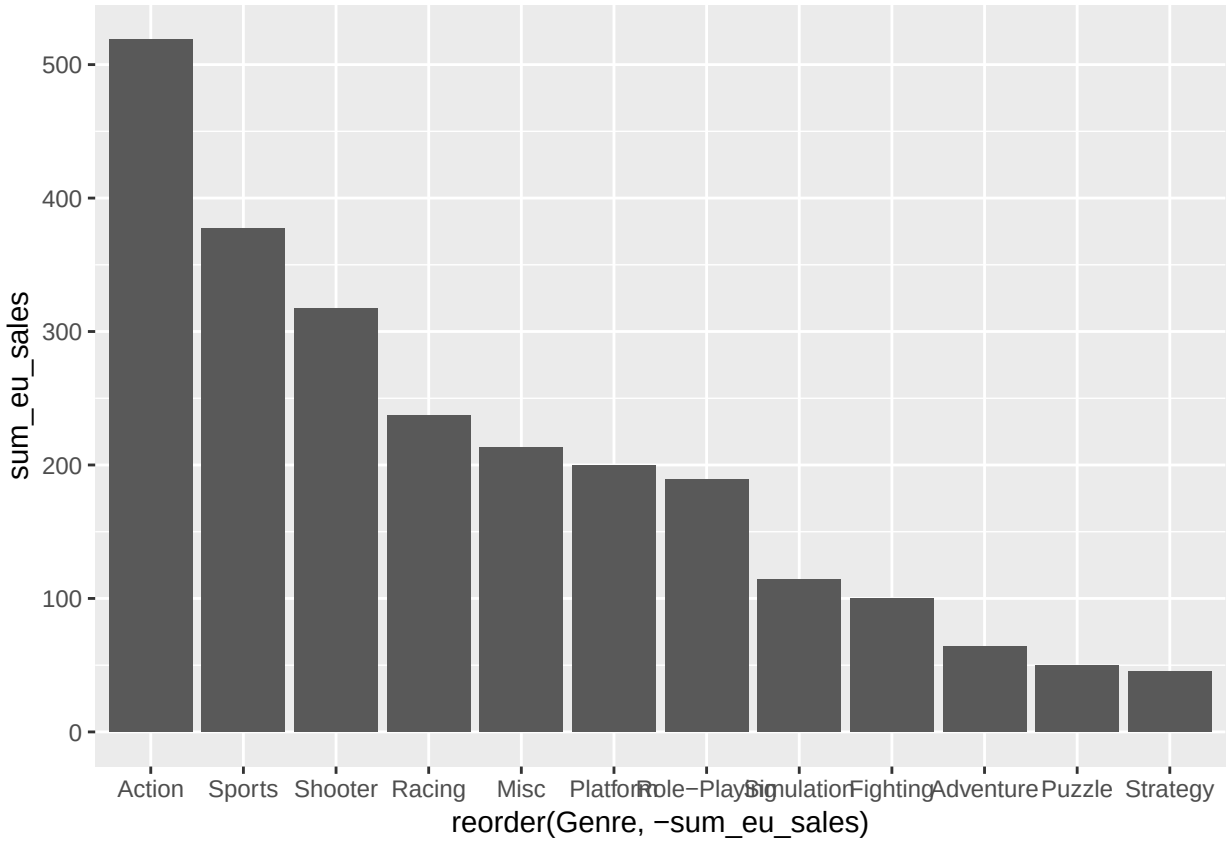
1-2. Most Popular Game Genres in Europe

```
genre_eu <- games %>%
  group_by(Genre) %>%
  summarise(sum_eu_sales = round(sum(EU_Sales), 0)) %>%
  arrange(desc(sum_eu_sales)) %>%
  mutate(eu_rank = dense_rank(desc(sum_eu_sales))) %>%
  head(12)

genre_eu
```

```
## # A tibble: 12 x 3
##    Genre        sum_eu_sales eu_rank
##    <fct>               <dbl>   <int>
##  1 Action                519       1
##  2 Sports                377       2
##  3 Shooter               317       3
##  4 Racing                237       4
##  5 Misc                  213       5
##  6 Platform              200       6
##  7 Role-Playing          189       7
##  8 Simulation            114       8
##  9 Fighting              100       9
## 10 Adventure              64      10
## 11 Puzzle                 50      11
## 12 Strategy               45      12
```

```
ggplot(data=genre_eu, aes(x=reorder(Genre, -sum_eu_sales), y=sum_eu_sales)) +
  geom_col()
```



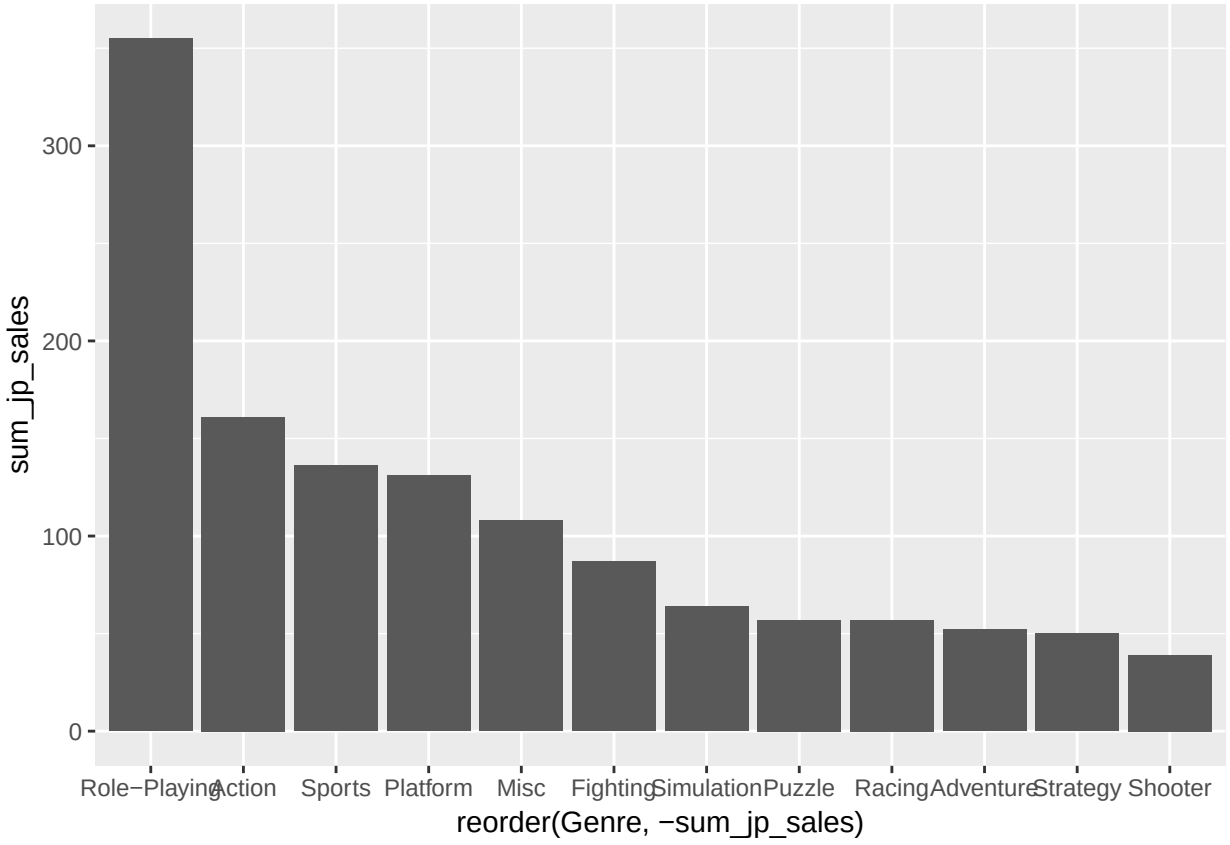
1-3. Most Popular Game Genres in Japan

```
genre_jp <- games %>%
  group_by(Genre) %>%
  summarise(sum_jp_sales = round(sum(JP_Sales), 0)) %>%
  arrange(desc(sum_jp_sales)) %>%
  mutate(jp_rank = dense_rank(desc(sum_jp_sales))) %>%
  head(12)

genre_jp
```

```
## # A tibble: 12 x 3
##    Genre        sum_jp_sales jp_rank
##    <fct>               <dbl>   <int>
##  1 Role-Playing          355       1
##  2 Action                161       2
##  3 Sports                136       3
##  4 Platform              131       4
##  5 Misc                  108       5
##  6 Fighting               87       6
##  7 Simulation             64       7
##  8 Puzzle                 57       8
##  9 Racing                 57       8
## 10 Adventure              52       9
## 11 Strategy               50      10
## 12 Shooter                39      11
```

```r
ggplot(data=genre_jp, aes(x=reorder(Genre, -sum_jp_sales), y=sum_jp_sales)) +
  geom_col()
```



1-4. Overall Rank

```r
popular_genre <- left_join(genre_na, genre_eu, by="Genre")
popular_genre <- left_join(popular_genre, genre_jp, by="Genre")
popular_genre
```

```
## # A tibble: 12 x 7
##    Genre     sum_na_sales na_rank sum_eu_sales eu_rank sum_jp_sales jp_rank
##    <fct>            <dbl>   <int>        <dbl>   <int>        <dbl>   <int>
##  1 Action             879       1          519       1          161       2
##  2 Sports             684       2          377       2          136       3
##  3 Shooter            592       3          317       3           39      11
##  4 Platform           446       4          200       6          131       4
##  5 Misc               407       5          213       5          108       5
##  6 Racing             359       6          237       4           57       8
##  7 Role-Pla~          331       7          189       7          355       1
##  8 Fighting           223       8          100       9           87       6
##  9 Simulati~          182       9          114       8           64       7
## 10 Puzzle             123      10           50      11           57       8
## 11 Adventure          105      11           64      10           52       9
## 12 Strategy            69      12           45      12           50      10
```

```
popular_genre <- popular_genre %>%
  select(Genre, na_rank, eu_rank, jp_rank)
popular_genre
```

```
## # A tibble: 12 x 4
##    Genre        na_rank eu_rank jp_rank
##    <fct>          <int>   <int>   <int>
##  1 Action             1       1       2
##  2 Sports             2       2       3
##  3 Shooter            3       3      11
##  4 Platform           4       6       4
##  5 Misc               5       5       5
##  6 Racing             6       4       8
##  7 Role-Playing       7       7       1
##  8 Fighting           8       9       6
##  9 Simulation         9       8       7
## 10 Puzzle            10      11       8
## 11 Adventure         11      10       9
## 12 Strategy          12      12      10
```

```
popular_genre <- popular_genre %>%
  mutate(sales_rank = dense_rank(na_rank+eu_rank+jp_rank)) %>%
  arrange(sales_rank)
popular_genre
```

```
## # A tibble: 12 x 5
##    Genre        na_rank eu_rank jp_rank sales_rank
##    <fct>          <int>   <int>   <int>      <int>
##  1 Action             1       1       2          1
##  2 Sports             2       2       3          2
##  3 Platform           4       6       4          3
##  4 Misc               5       5       5          4
##  5 Role-Playing       7       7       1          4
##  6 Shooter            3       3      11          5
##  7 Racing             6       4       8          6
##  8 Fighting           8       9       6          7
##  9 Simulation         9       8       7          8
## 10 Puzzle            10      11       8          9
## 11 Adventure         11      10       9         10
## 12 Strategy          12      12      10         11
```

- Role-Playing, Action, Sports, Platform, Misc, Fighting, Simulation, Racing appeared in top 10 lists of all three regions.
- NA and EU's top 3 were Action-Sports-Shooter, and Role-Playing came 7 in both.
- However in JP, Role-Playing came first and its difference to second place is greater than double in size, and Shooter listed 11.
- NA and Eu have quite similar ranks, but totally different in JP.

2. Critic Score and User Score of Genres

```
class(games$Critic_Score)
```

```
## [1] "integer"
```

```
table(is.na(games$Critic_Score))
```

```
##
## FALSE  TRUE
##  8137  8582
```

There are 8582 rows with NA citric score. Creating a default value or an average value do not really make sense here, so I will just get the rows with no NA value. And to increase the accuracy, I will consider the rows with

- Greater than or equal to 5 critic reviews
- Greater than or equal to 20 user reviews

```
critic_genre <- games %>%
  filter(!is.na(Critic_Score)) %>%
  filter(Critic_Count >= 5 & User_Count >= 20) %>%
  group_by(Genre) %>%
  summarise(mean_critic_score = mean(Critic_Score))

head(critic_genre)
```

```
## # A tibble: 6 x 2
##   Genre     mean_critic_score
##   <fct>                 <dbl>
## 1 Action                 72.3
## 2 Adventure              73.5
## 3 Fighting               73.9
## 4 Misc                   74.2
## 5 Platform               75
## 6 Puzzle                 78.7
```

```
class(games$User_Score)
```

```
## [1] "factor"
```

```
table(is.na(games$User_Score))
```

```
##
## FALSE
## 16719
```

```
games$User_Score <- as.integer(games$User_Score)
class(games$User_Score)
```

```
## [1] "integer"
```

```
user_genre <- games %>%
  filter(!is.na(User_Score)) %>%
  filter(Critic_Count >= 5 & User_Count >= 20) %>%
  group_by(Genre) %>%
  summarise(mean_user_score = mean(User_Score))

head(user_genre)
```

```
## # A tibble: 6 x 2
##   Genre     mean_user_score
##   <fct>               <dbl>
## 1 Action               72.1
```

```
## 2 Adventure            76.9
## 3 Fighting             74.5
## 4 Misc                 72.2
## 5 Platform             75.9
## 6 Puzzle               78.4
```

```
genre_score <- left_join(critic_genre, user_genre, by="Genre")
genre_score
```

```
## # A tibble: 12 x 3
##    Genre        mean_critic_score mean_user_score
##    <fct>                    <dbl>           <dbl>
##  1 Action                    72.3            72.1
##  2 Adventure                 73.5            76.9
##  3 Fighting                  73.9            74.5
##  4 Misc                      74.2            72.2
##  5 Platform                  75              75.9
##  6 Puzzle                    78.7            78.4
##  7 Racing                    75.7            71.6
##  8 Role-Playing              75.3            75.8
##  9 Shooter                   73.9            70.3
## 10 Simulation                73.6            72.4
## 11 Sports                    77.4            68.6
## 12 Strategy                  76.2            72.4
```

- Genre with highest critic score is Puzzle
- Genre with highest user score is Puzzle, too.
- This is unpredictable since Puzzle came 10 in NA, 11 in EU and 8 in JP at the previous sales rank section.
- Action, which has the highest sales amount, has fairly low scores compared to the others.
- This tells me that sales amount and score do not coorelate.

```
genre_score <- genre_score %>%
  mutate(tot_score = (mean_critic_score + mean_user_score)/2) %>%
  mutate(score_rank = dense_rank(desc(tot_score))) %>%
  arrange(score_rank)
genre_score
```

```
## # A tibble: 12 x 5
##    Genre        mean_critic_score mean_user_score tot_score score_rank
##    <fct>                    <dbl>           <dbl>     <dbl>      <int>
##  1 Puzzle                    78.7            78.4      78.5          1
##  2 Role-Playing              75.3            75.8      75.5          2
##  3 Platform                  75              75.9      75.5          3
##  4 Adventure                 73.5            76.9      75.2          4
##  5 Strategy                  76.2            72.4      74.3          5
##  6 Fighting                  73.9            74.5      74.2          6
##  7 Racing                    75.7            71.6      73.6          7
##  8 Misc                      74.2            72.2      73.2          8
##  9 Sports                    77.4            68.6      73.0          9
## 10 Simulation                73.6            72.4      73.0         10
## 11 Action                    72.3            72.1      72.2         11
## 12 Shooter                   73.9            70.3      72.1         12
```

```
genre_sales_score_rank <- left_join(popular_genre, genre_score, by="Genre")
genre_sales_score_rank <- genre_sales_score_rank %>%
```

```
  select(Genre, sales_rank, score_rank)
genre_sales_score_rank
```

```
## # A tibble: 12 x 3
##    Genre        sales_rank score_rank
##    <fct>             <int>      <int>
##  1 Action                1         11
##  2 Sports                2          9
##  3 Platform              3          3
##  4 Misc                  4          8
##  5 Role-Playing          4          2
##  6 Shooter               5         12
##  7 Racing                6          7
##  8 Fighting              7          6
##  9 Simulation            8         10
## 10 Puzzle                9          1
## 11 Adventure            10          4
## 12 Strategy             11          5
```

- Overall, Platform and Role-Playing games have the highest rank (sales + score).
- Simulation games have the lowest rank.
- Action, Sports and Shooter games are the most popular games in NA and EU, however their reviews are not as good as their sales amount(11, 9, 12 respectively).

3. Publishers with Good Reputation

Amount of published games of certain companies does not really mean the reason to buy their game, but at least "amount" is a distinct number that does not change. This can show the company's history, skills and effort to create a new game.

```
publisher <- games %>%
  group_by(Publisher) %>%
  summarise(n=n()) %>%
  arrange(desc(n)) %>%
  head(10)
publisher
```

```
## # A tibble: 10 x 2
##    Publisher                       n
##    <fct>                       <int>
##  1 Electronic Arts              1356
##  2 Activision                    985
##  3 Namco Bandai Games            939
##  4 Ubisoft                       933
##  5 Konami Digital Entertainment  834
##  6 THQ                           715
##  7 Nintendo                      706
##  8 Sony Computer Entertainment   687
##  9 Sega                          638
## 10 Take-Two Interactive          422
```

- These are the top 10 games with highest amount of released games.

```
publisher_genre <- games %>%
  filter(Publisher %in% c("Electronic Arts",
                          "Activision",
```

```
                          "Namco Bandai Games",
                          "Ubisoft",
                          "Konami Digital Entertainment",
                          "THQ",
                          "Nintendo",
                          "Sony Computer Entertainment",
                          "Sega",
                          "Take-Two Interactive")) %>%
  filter(Genre %in% c("Platform", "Role-Playing")) %>%
  filter(Critic_Score > 75 & User_Score >75) %>%
  arrange(desc(Global_Sales)) %>%
  head(10)
publisher_genre
```

```
##                          Name Platform Year_of_Release        Genre
## 1        New Super Mario Bros.       DS            2006     Platform
## 2    New Super Mario Bros. Wii      Wii            2009     Platform
## 3           Super Mario Galaxy      Wii            2007     Platform
## 4          Super Mario 3D Land      3DS            2011     Platform
## 5            Final Fantasy VII       PS            1997 Role-Playing
## 6              Final Fantasy X      PS2            2001 Role-Playing
## 7    PokÃ©mon Platinum Version       DS            2008 Role-Playing
## 8         Super Mario Galaxy 2      Wii            2010     Platform
## 9    Crash Bandicoot 3: Warped       PS            1998     Platform
## 10 Donkey Kong Country Returns      Wii            2010     Platform
##                      Publisher NA_Sales EU_Sales JP_Sales Other_Sales
## 1                     Nintendo    11.28     9.14     6.50        2.88
## 2                     Nintendo    14.44     6.94     4.70        2.24
## 3                     Nintendo     6.06     3.35     1.20        0.74
## 4                     Nintendo     4.89     3.00     2.14        0.78
## 5  Sony Computer Entertainment     3.01     2.47     3.28        0.96
## 6  Sony Computer Entertainment     2.91     2.07     2.73        0.33
## 7                     Nintendo     2.76     1.72     2.69        0.54
## 8                     Nintendo     3.56     2.35     0.98        0.62
## 9  Sony Computer Entertainment     3.68     1.75     1.42        0.28
## 10                    Nintendo     3.17     1.79     1.03        0.46
##    Global_Sales Critic_Score Critic_Count User_Score User_Count
## 1         29.80           89           65         84        431
## 2         28.32           87           80         83        594
## 3         11.35           97           73         88       2147
## 4         10.81           90           82         83        921
## 5          9.72           92           20         91       1282
## 6          8.05           92           53         86       1056
## 7          7.72           83           46         84        203
## 8          7.51           97           87         90       1854
## 9          7.13           91           12         88        432
## 10         6.44           87           77         85        368
##              Developer Rating
## 1             Nintendo      E
## 2             Nintendo      E
## 3             Nintendo      E
## 4             Nintendo      E
## 5           SquareSoft      T
## 6           SquareSoft      T
```

```
## 7          Game Freak      E
## 8  Nintendo EAD Tokyo      E
## 9         Naughty Dog      E
## 10      Retro Studios      E
```

4. Conclusion

Applying certain aspects that I have found above; - Either Role-Playing or Platform - Published by the publishers with highest product number (top 10) - Critic and User scores greater than 75

Super Mario series are the best games to have a try. One curiosity is that, Action and Sports games have such a low review score but their sales amount are high in every region. My hypothesis is that users and critics are becoming more harsh on Action and Sports games since they have high expactation against those games.