

choose your own project

John Kingsley

2019/12/24

INTRODUCTION Video games are widely played all over the world and the industries in this area of business have tremendously grown in the past decades. Video games are electronic designed algorithm that are incorporated into a computing device which includes computers (either desktop or laptop), mobile phone or a gaming console. Video games are subdivided into mobile games and computers, which solely dependent on the platforms. In the early 80s and late 70s, there were two major markets for video games, the home and the arcade markets. The arcade market generated approximately 8 billion USD more than pop music as at 1982 drawing huge attention from investors. The video games have then grown tremendously and the games enjoyed all over the world by all class of people irrespective of their age. More so, some video games have come and gone while others still in the industry are developing modern games with the needs of the people and with the growing technologies VIDEO GAMES INDUSTRY PROJECTION In 2016, The market research conducted by the Newzoo firm showed that it predicts the global game market to grow up to \$99.6 billion which was about 8.4 percent when compared to that of the previous year. In this year, the firm also predicted that mobile games will experience more sales for the first time over the console games industry with 21.3 percent growth when compared to previous year. TRENDS IN THE INDUSTRY OVER THE YEARS 1. Sales of games are dominant in the North America region 2.puzzled games has obviously declined in the game popularity while Action and adventure games have experienced positive growth.

PROJECT OVERVIEW The main aim of the project is to show relationship in sales Other aims includes a. Sales at different regions of the world b. Popular genre at global sales c. Popular Publishers at global sales d. The year with the highest number of sales. a. to see if there are relationships between sales of the regions b. to see if there are relationship between sales and genres in the different regions of the world This we can achieve through data visualization after that we can further do the following data analysis, to note observable changes in comparing the genre of a game and the platform which they are released in, to ascertain if there are any relationship between sales in the regions, genres and sales.

VARIABLES DEFINITION 1.Name: Name of the video game 2.Platform: Platform on which the game was released or is playable 3.Year: Year in which the game was released 4.Genre: Genre the game belongs to 5.Publisher: Name of the publisher who created the game 6.NA_Sales: Sales in North America 7.EU_Sales: Sales in Europe 8. JP_Sales: Sales in Japan 9. Other_Sales: Sales in other countries 10. Global_Sales: Global Sales

DATA The data for the project was obtained from (https://www.kaggle.com/rush4ratio/video-game-sales-with-ratings/data#Video_Games_Sales_as_at_22_Dec_2016.csv) (https://www.kaggle.com/rush4ratio/video-game-sales-with-ratings/data#Video_Games_Sales_as_at_22_Dec_2016.csv) which was already in it tidy state and available for machine learning analysis. The data provided insight on the trends of video game industry for over 30 years in the different regions of the world and also at global perspective.

EXECUIVE SUMMARY The analysis offered me great insight into the details of the video game industry which are widely played and enjoyed by all regions of the world regardless of age. The analysis revealed that there is a relationship between regional sales and global. That the increase or decrease in regional sales will proportional influence global sales. And that North America regions are the highest consumers of video games followed by European Union. Similarly, Shooter, Sport and Action genres have maintained top spot in the video game industry over the years. And that DS, PS, XBOX and Wii are in the top spots of the video game industry.

METHDOLOGY, ANALYSIS AND RESULT #load packages

```
if(!require(tidyverse)) install.packages("tidyverse", repos = "http://cran.us.r-project.org")
```

```
## Loading required package: tidyverse
```

```
## — Attaching packages ————— tidyverse 1.3.0 —
```

```
## <U+2713> ggplot2 3.2.1      <U+2713> purrr   0.3.3
## <U+2713> tibble  2.1.3      <U+2713> dplyr   0.8.3
## <U+2713> tidyr   1.0.0      <U+2713> stringr 1.4.0
## <U+2713> readr   1.3.1      <U+2713>forcats 0.4.0
```

```
## — Conflicts ————— tidyverse_conflicts() —
```

```
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()   masks stats::lag()
```

```
if(!require(caret)) install.packages("caret", repos = "http://cran.us.r-project.org")
```

```
## Loading required package: caret
```

```
## Loading required package: lattice
```

```
## 
## Attaching package: 'caret'
```

```
## The following object is masked from 'package:purrr':  
##  
##     lift
```

```
if(!require(data.table)) install.packages("data.table", repos = "http://cran.us.r-project.org")
```

```
## Loading required package: data.table
```

```
##  
## Attaching package: 'data.table'
```

```
## The following objects are masked from 'package:dplyr':  
##  
##     between, first, last
```

```
## The following object is masked from 'package:purrr':  
##  
##     transpose
```

```
library(viridis)
```

```
## Loading required package: viridisLite
```

```
library(wordcloud)
```

```
## Loading required package: RColorBrewer
```

```
library(RColorBrewer)  
library(magrittr)
```

```
##  
## Attaching package: 'magrittr'
```

```
## The following object is masked from 'package:purrr':  
##  
##     set_names
```

```
## The following object is masked from 'package:tidyR':  
##  
##     extract
```

```
library(lubridate)
```

```
##  
## Attaching package: 'lubridate'
```

```
## The following objects are masked from 'package:data.table':  
##  
##     hour, isoweek, mday, minute, month, quarter, second, wday, week,  
##     yday, year
```

```
## The following object is masked from 'package:base':  
##  
##     date
```

```
library(RPostgreSQL)
```

```
## Loading required package: DBI
```

```
library(plotly)
```

```
##  
## Attaching package: 'plotly'
```

```
## The following object is masked from 'package:ggplot2':  
##  
##     last_plot
```

```
## The following object is masked from 'package:stats':  
##  
##     filter
```

```
## The following object is masked from 'package:graphics':  
##  
##     layout
```

```
library(jsonlite)
```

```
##  
## Attaching package: 'jsonlite'
```

```
## The following object is masked from 'package:purrr':  
##  
##     flatten
```

```
library(htmltools)  
library(glmnet)
```

```
## Loading required package: Matrix
```

```
##  
## Attaching package: 'Matrix'
```

```
## The following objects are masked from 'package:tidyverse':  
##  
##     expand, pack, unpack
```

```
## Loaded glmnet 3.0-2
```

```
library(epitools)  
library(lme4)  
library(sjPlot)
```

```
## #refugeeswelcome
```

```
library(pscl)
```

```
## Classes and Methods for R developed in the  
## Political Science Computational Laboratory  
## Department of Political Science  
## Stanford University  
## Simon Jackman  
## hurdle and zeroinfl functions by Achim Zeileis
```

```
#loading of dataset
```

```
VIDEOGS <- read.csv("D:/john/Videogs_2016.csv")
```

```
#data inspection
```

```
head(VIDEOGS)
```

```

##          Name Platform Year_of_Release      Genre Publisher
## 1      Wii Sports      Wii           2006      Sports  Nintendo
## 2 Super Mario Bros.     NES          1985   Platform  Nintendo
## 3      Mario Kart Wii      Wii           2008      Racing  Nintendo
## 4    Wii Sports Resort      Wii           2009      Sports  Nintendo
## 5 Pokemon Red/Pokemon Blue     GB          1996 Role-Playing  Nintendo
## 6        Tetris       GB          1989      Puzzle  Nintendo
##   NA_Sales EU_Sales JP_Sales Other_Sales Global_Sales Critic_Score Critic_Count
## 1    41.36    28.96    3.77      8.45     82.53        76         51
## 2    29.08    3.58    6.81      0.77     40.24       NA        NA
## 3    15.68   12.76    3.79      3.29     35.52        82         73
## 4    15.61   10.93    3.28      2.95     32.77        80         73
## 5    11.27    8.89   10.22      1.00     31.37       NA        NA
## 6    23.20    2.26    4.22      0.58     30.26       NA        NA
##   User_Score User_Count Developer Rating
## 1          8        322  Nintendo     E
## 2          NA
## 3        8.3        709  Nintendo     E
## 4          8        192  Nintendo     E
## 5          NA
## 6          NA

```

```
glimpse(VIDEOGS)
```

```

## # Observations: 16,719
## # Variables: 16
## $ Name          <fct> Wii Sports, Super Mario Bros., Mario Kart Wii, Wii Sp...
## $ Platform       <fct> Wii, NES, Wii, Wii, GB, GB, DS, Wii, NES, DS, DS...
## $ Year_of_Release <fct> 2006, 1985, 2008, 2009, 1996, 1989, 2006, 2006, 2009, ...
## $ Genre          <fct> Sports, Platform, Racing, Sports, Role-Playing, Puzzl...
## $ Publisher      <fct> Nintendo, Nintendo, Nintendo, Nintendo, Nin...
## $ NA_Sales        <dbl> 41.36, 29.08, 15.68, 15.61, 11.27, 23.20, 11.28, 13.9...
## $ EU_Sales        <dbl> 28.96, 3.58, 12.76, 10.93, 8.89, 2.26, 9.14, 9.18, 6...
## $ JP_Sales        <dbl> 3.77, 6.81, 3.79, 3.28, 10.22, 4.22, 6.50, 2.93, 4.70...
## $ Other_Sales     <dbl> 8.45, 0.77, 3.29, 2.95, 1.00, 0.58, 2.88, 2.84, 2.24, ...
## $ Global_Sales    <dbl> 82.53, 40.24, 35.52, 32.77, 31.37, 30.26, 29.80, 28.9...
## $ Critic_Score    <int> 76, NA, 82, 80, NA, NA, 89, 58, 87, NA, NA, 91, NA, 8...
## $ Critic_Count    <int> 51, NA, 73, 73, NA, NA, 65, 41, 80, NA, NA, 64, NA, 6...
## $ User_Score       <fct> 8, , 8.3, 8, , , 8.5, 6.6, 8.4, , , 8.6, , 7.7, 6.3, ...
## $ User_Count       <int> 322, NA, 709, 192, NA, NA, 431, 129, 594, NA, NA, 464...
## $ Developer        <fct> Nintendo, , Nintendo, Nintendo, , , Nintendo, Nintend...
## $ Rating          <fct> E, , E, E, , E, E, , E, E, M, M, , E, ...

```

```
summary(VIDEOGS)
```

```

##          Name      Platform Year_of_Release
## Need for Speed: Most Wanted: 12   PS2     :2161  2008   :1427
## FIFA 14                 : 9    DS      :2152  2009   :1426
## LEGO Marvel Super Heroes : 9    PS3     :1330  2010   :1254
## Madden NFL 07            : 9    Wii     :1320  2007   :1196
## Ratatouille              : 9    X360    :1262  2011   :1136
## Angry Birds Star Wars   : 8    PSP     :1208  2006   :1006
## (Other)                  :16663 (Other):7286 (Other):9274
##          Genre      Publisher NA_Sales
## Action      :3370  Electronic Arts       : 1356 Min.   : 0.0000
## Sports      :2348  Activision           :  985 1st Qu.: 0.0000
## Misc        :1750  Namco Bandai Games      :  939 Median  : 0.0800
## Role-Playing:1500 Ubisoft             :  933 Mean    : 0.2633
## Shooter     :1323  Konami Digital Entertainment: 834 3rd Qu.: 0.2400
## Adventure   :1301  THQ                :  715 Max.   :41.3600
## (Other)      :5127  (Other)            :10957
##          EU_Sales   JP_Sales   Other_Sales Global_Sales
## Min.   : 0.000  Min.   :0.00000  Min.   : 0.00000  Min.   : 0.0100
## 1st Qu.: 0.000  1st Qu.:0.00000  1st Qu.: 0.00000  1st Qu.: 0.0600
## Median : 0.020  Median :0.00000  Median : 0.01000  Median : 0.1700
## Mean   : 0.145  Mean   : 0.07759  Mean   : 0.04734  Mean   : 0.5336
## 3rd Qu.: 0.110  3rd Qu.:0.04000  3rd Qu.: 0.03000  3rd Qu.: 0.4700
## Max.   :28.960  Max.   :10.22000  Max.   :10.57000  Max.   :82.5300
## NA's   :8582    NA's   :8582    (Other):6440    NA's   :9129
##          Developer   Rating
## :6623        :6769
## Ubisoft     : 204   E     :3991
## EA Sports    172   T     :2961
## EA Canada   167   M     :1563
## Konami      : 162   E10+  :1420
## Capcom      : 139   EC    :  8
## (Other)     :9252  (Other):  7

```

```
names(VIDEOGS)
```

```

## [1] "Name"          "Platform"      "Year_of_Release" "Genre"
## [5] "Publisher"     "NA_Sales"      "EU_Sales"       "JP_Sales"
## [9] "Other_Sales"    "Global_Sales"   "Critic_Score"   "Critic_Count"
## [13] "User_Score"     "User_Count"    "Developer"      "Rating"

```

```
dim(VIDEOGS)
```

```
## [1] 16719    16
```

```
#class of some variables
```

```
class(VIDEOGS$Name)
```

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

```
class(VIDEOGS$Platform)
```

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

```
class(VIDEOGS$Publisher)
```

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

```
class(VIDEOGS$Year_of_Release)
```

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

```
class(VIDEOGS$Genre)
```

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

```
#load package
```

```
library(psych)
```

```
##  
## Attaching package: 'psych'
```

```
## The following objects are masked from 'package:ggplot2':
```

```
##  
##     %+%, alpha
```

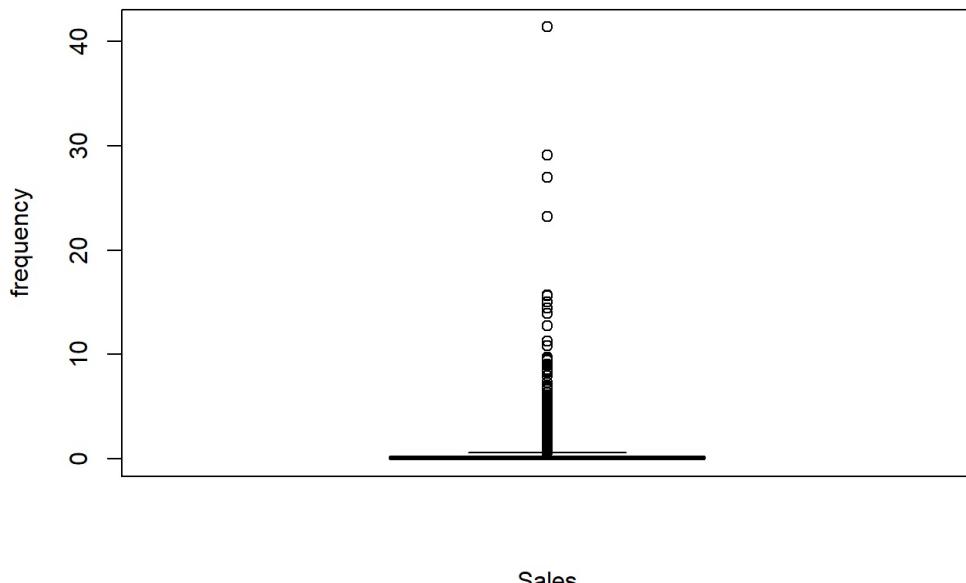
```
describe(VIDEOGS)
```

| | vars | n | mean | sd | median | trimmed | mad | min |
|---------------------|------|----------|----------|---------|---------|----------|---------|-------|
| ## Name* | 1 | 16719 | 5830.14 | 3345.05 | 5898.00 | 5844.17 | 4306.95 | 1.00 |
| ## Platform* | 2 | 16719 | 18.74 | 8.29 | 19.00 | 18.70 | 10.38 | 1.00 |
| ## Year_of_Release* | 3 | 16719 | 27.71 | 6.08 | 29.00 | 28.07 | 4.45 | 1.00 |
| ## Genre* | 4 | 16719 | 7.76 | 4.41 | 8.00 | 7.65 | 5.93 | 1.00 |
| ## Publisher* | 5 | 16719 | 302.15 | 183.01 | 332.00 | 306.78 | 274.28 | 1.00 |
| ## NA_Sales | 6 | 16719 | 0.26 | 0.81 | 0.08 | 0.13 | 0.12 | 0.00 |
| ## EU_Sales | 7 | 16719 | 0.15 | 0.50 | 0.02 | 0.06 | 0.03 | 0.00 |
| ## JP_Sales | 8 | 16719 | 0.08 | 0.31 | 0.00 | 0.02 | 0.00 | 0.00 |
| ## Other_Sales | 9 | 16719 | 0.05 | 0.19 | 0.01 | 0.02 | 0.01 | 0.00 |
| ## Global_Sales | 10 | 16717 | 0.53 | 1.55 | 0.17 | 0.27 | 0.21 | 0.01 |
| ## Critic_Score | 11 | 8137 | 68.97 | 13.94 | 71.00 | 69.88 | 13.34 | 13.00 |
| ## Critic_Count | 12 | 8137 | 26.36 | 18.98 | 21.00 | 23.85 | 16.31 | 3.00 |
| ## User_Score* | 13 | 16719 | 46.36 | 39.47 | 61.00 | 45.70 | 53.37 | 1.00 |
| ## User_Count | 14 | 7590 | 162.23 | 561.28 | 24.00 | 49.96 | 26.69 | 4.00 |
| ## Developer* | 15 | 16719 | 514.29 | 570.72 | 302.00 | 445.28 | 446.26 | 1.00 |
| ## Rating* | 16 | 16719 | 3.71 | 3.01 | 3.00 | 3.39 | 2.97 | 1.00 |
| | | | max | range | skew | kurtosis | se | |
| ## Name* | | 11563.00 | 11562.00 | -0.03 | -1.21 | 25.87 | | |
| ## Platform* | | 33.00 | 32.00 | -0.05 | -0.99 | 0.06 | | |
| ## Year_of_Release* | | 41.00 | 40.00 | -0.80 | 1.62 | 0.05 | | |
| ## Genre* | | 15.00 | 14.00 | 0.08 | -1.35 | 0.03 | | |
| ## Publisher* | | 583.00 | 582.00 | -0.15 | -1.39 | 1.42 | | |
| ## NA_Sales | | 41.36 | 41.36 | 18.77 | 648.43 | 0.01 | | |
| ## EU_Sales | | 28.96 | 28.96 | 18.85 | 755.36 | 0.00 | | |
| ## JP_Sales | | 10.22 | 10.22 | 11.21 | 194.23 | 0.00 | | |
| ## Other_Sales | | 10.57 | 10.57 | 24.58 | 1054.69 | 0.00 | | |
| ## Global_Sales | | 82.53 | 82.52 | 17.37 | 603.78 | 0.01 | | |
| ## Critic_Score | | 98.00 | 85.00 | -0.61 | 0.14 | 0.15 | | |
| ## Critic_Count | | 113.00 | 110.00 | 1.15 | 1.03 | 0.21 | | |
| ## User_Score* | | 97.00 | 96.00 | -0.11 | -1.73 | 0.31 | | |
| ## User_Count | | 10665.00 | 10661.00 | 9.03 | 112.41 | 6.44 | | |
| ## Developer* | | 1697.00 | 1696.00 | 0.69 | -1.00 | 4.41 | | |
| ## Rating* | | 9.00 | 8.00 | 0.78 | -0.92 | 0.02 | | |

```
#Sales trend around the regions of the world according to platforms, genres, names and publishers. #Sales in North America
```

```
boxplot(VIDEOGS$NA_Sales, main="Sales in North America", xlab="Sales", ylab="frequency", vertical = TRUE)
```

Sales in North America



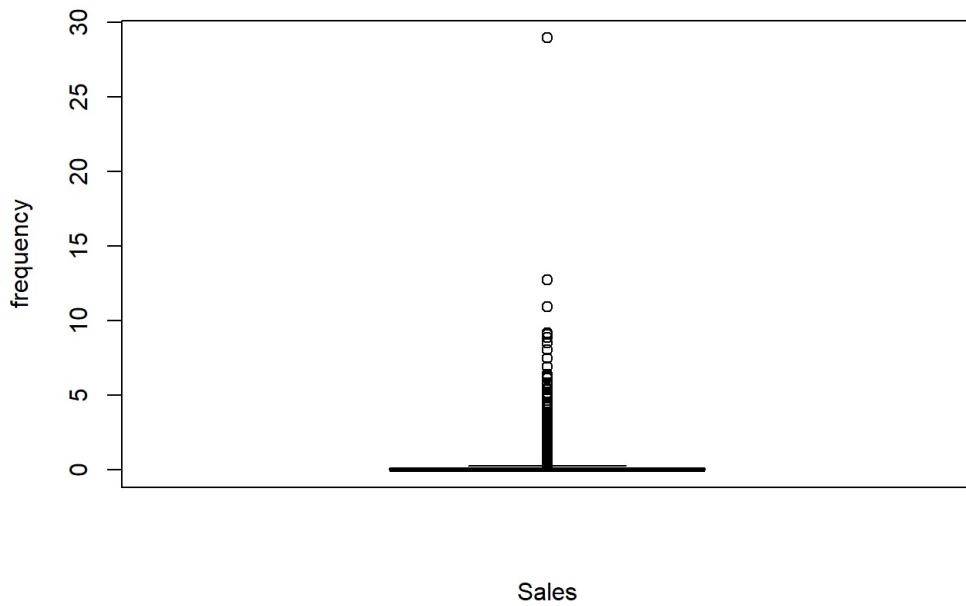
```
summary(VIDEOGS$NA_Sales)
```

```
##      Min. 1st Qu. Median      Mean 3rd Qu.      Max.
## 0.0000 0.0000 0.0800 0.2633 0.2400 41.3600
```

```
#sales in Europe
```

```
boxplot(VIDEOGS$EU_Sales, main="Sales in Europe", xlab="Sales", ylab="frequency", vertical = TRUE)
```

Sales in Europe



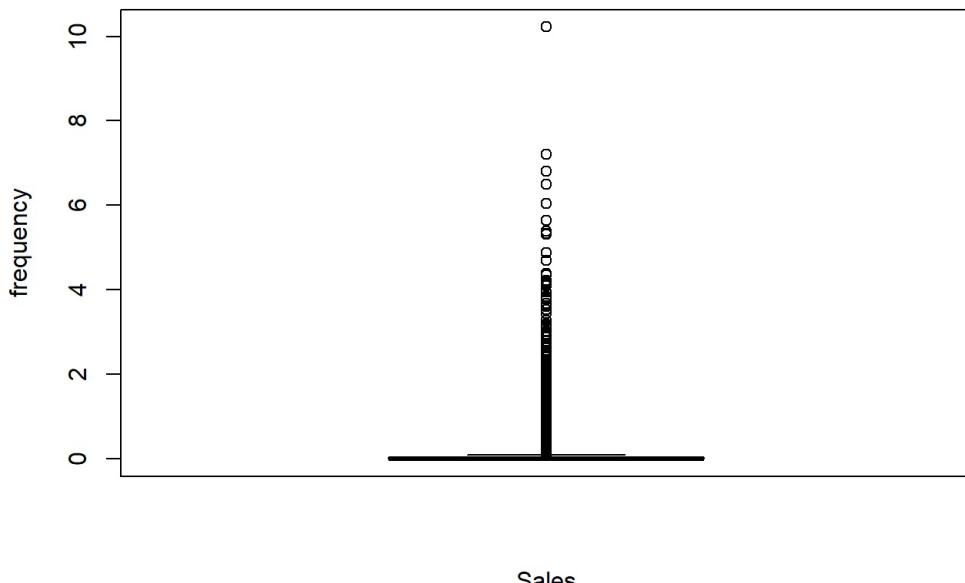
```
summary(VIDEOGS$EU_Sales)
```

```
##      Min. 1st Qu. Median      Mean 3rd Qu.      Max.
## 0.000 0.000 0.020 0.145 0.110 28.960
```

```
#sales in Japan
```

```
boxplot(VIDEOGS$JP_Sales, main="Sales in Japan", xlab="Sales", ylab="frequency", vertical = TRUE)
```

Sales in Japan



Sales

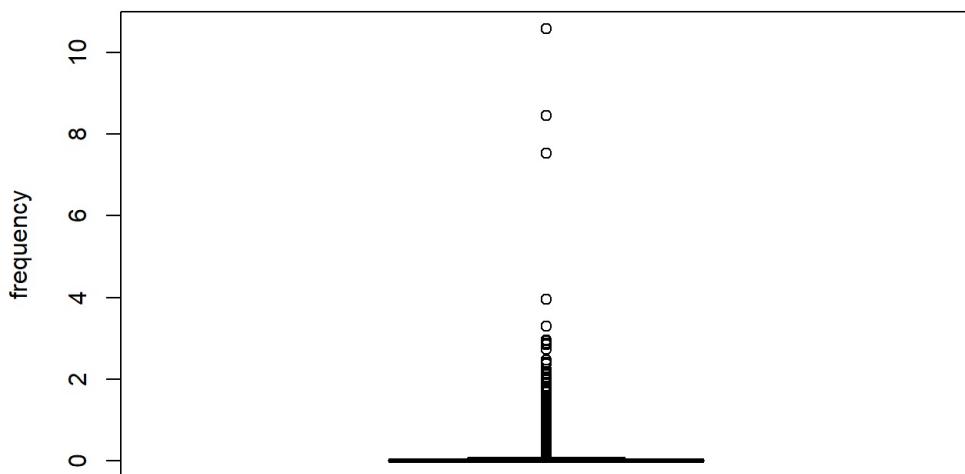
```
summary(VIDEOGS$JP_Sales)
```

```
##      Min. 1st Qu. Median     Mean 3rd Qu.    Max.  
## 0.00000 0.00000 0.00000 0.07759 0.04000 10.22000
```

#other sales

```
boxplot(VIDEOGS$Other_Sales, main="Sales in Japan", xlab="Sales", ylab="frequency", vertical = TRUE)
```

Sales in Japan



Sales

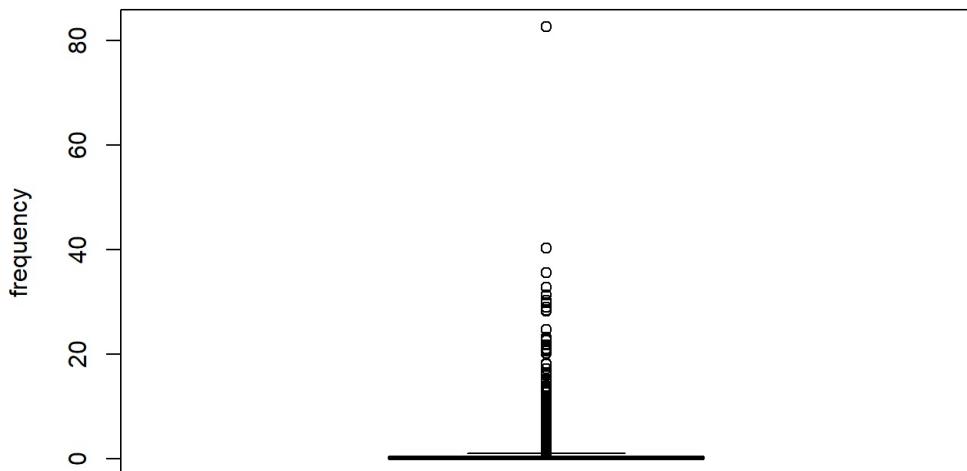
```
summary(VIDEOGS$JP_Sales)
```

```
##      Min. 1st Qu. Median     Mean 3rd Qu.    Max.  
## 0.00000 0.00000 0.00000 0.07759 0.04000 10.22000
```

#Global sales

```
boxplot(VIDEOGS$Global_Sales, main="Sales in Japan", xlab="Sales", ylab="frequency", vertical = TRUE)
```

Sales in Japan



Sales

```
summary(VIDEOGS$Global_Sales)
```

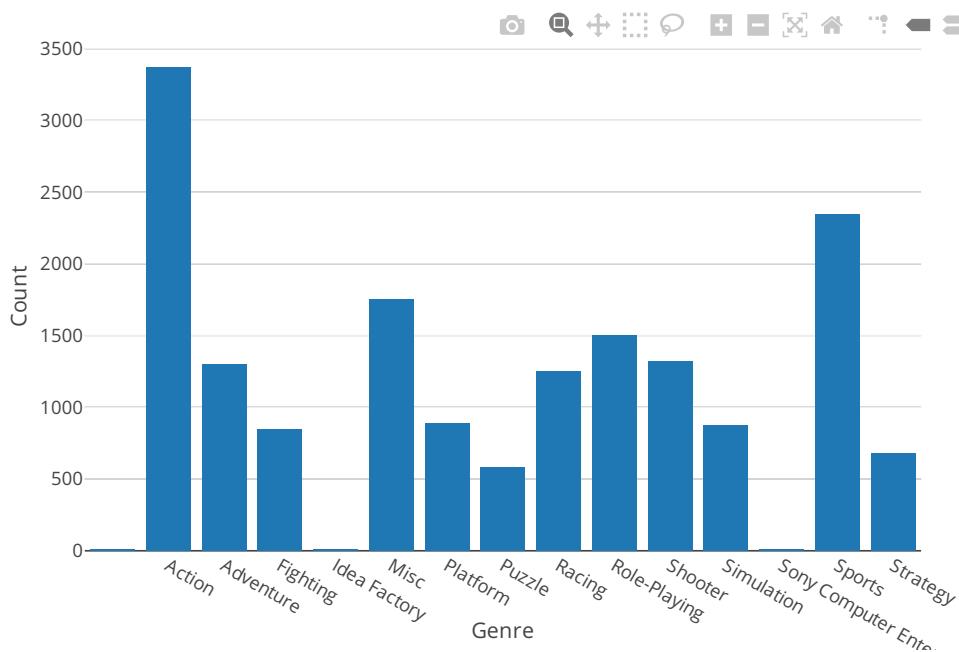
```
##      Min. 1st Qu. Median     Mean 3rd Qu.    Max.    NA's
## 0.0100 0.0600 0.1700 0.5336 0.4700 82.5300      2
```

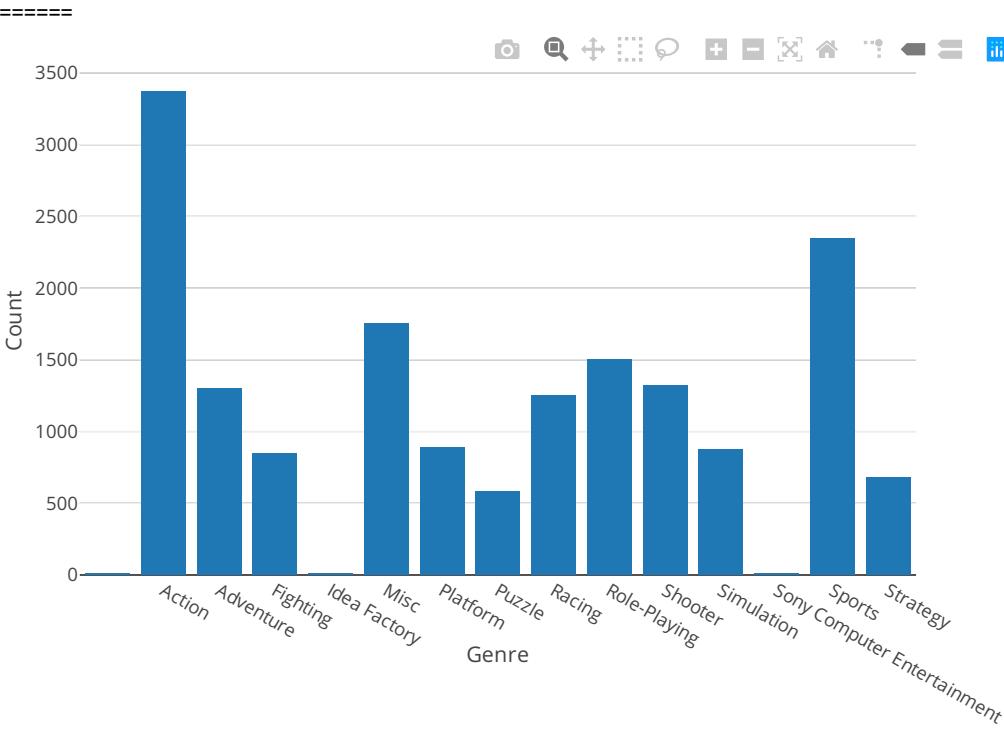
#Popular Genre by sales at global level

```
VIDEOGS %>%
  group_by(Genre) %>%
  summarise(Count = n()) %>%
  plot_ly(x = ~Genre,
          y = ~Count,
          type = "bar", col="black")
```

```
## Warning: 'bar' objects don't have these attributes: 'col'
## Valid attributes include:
## 'type', 'visible', 'showlegend', 'legendgroup', 'opacity', 'name', 'uid', 'ids', 'customdata', 'meta', 'selecte
dpoints', 'hoverinfo', 'hoverlabel', 'stream', 'transforms', 'uirevision', 'x', 'x0', 'dx', 'y', 'y0', 'dy', 'text
', 'hovertext', 'hovertemplate', 'textposition', 'insidetextanchor', 'textangle', 'textfont', 'insidetextfont', 'o
utsidetextfont', 'constrainttext', 'cliponaxis', 'orientation', 'base', 'offset', 'width', 'marker', 'offsetgroup',
'alignmentgroup', 'selected', 'unselected', 'r', 't', '_deprecated', 'error_x', 'error_y', 'xcalendar', 'ycalendar
', 'xaxis', 'yaxis', 'idssrc', 'customdatasrc', 'metasrc', 'hoverinfosrc', 'xsrc', 'ysrc', 'textsrc', 'hovertextr
src', 'hovertemplatesrc', 'textpositions', 'basesrc', 'offsetsrc', 'widthsrc', 'rsrc', 'tsrc', 'key', 'set', 'fram
e', 'transforms', '_isNestedKey', '_isSimpleKey', '_isGraticule', '_bbox'
```

<<<<< HEAD



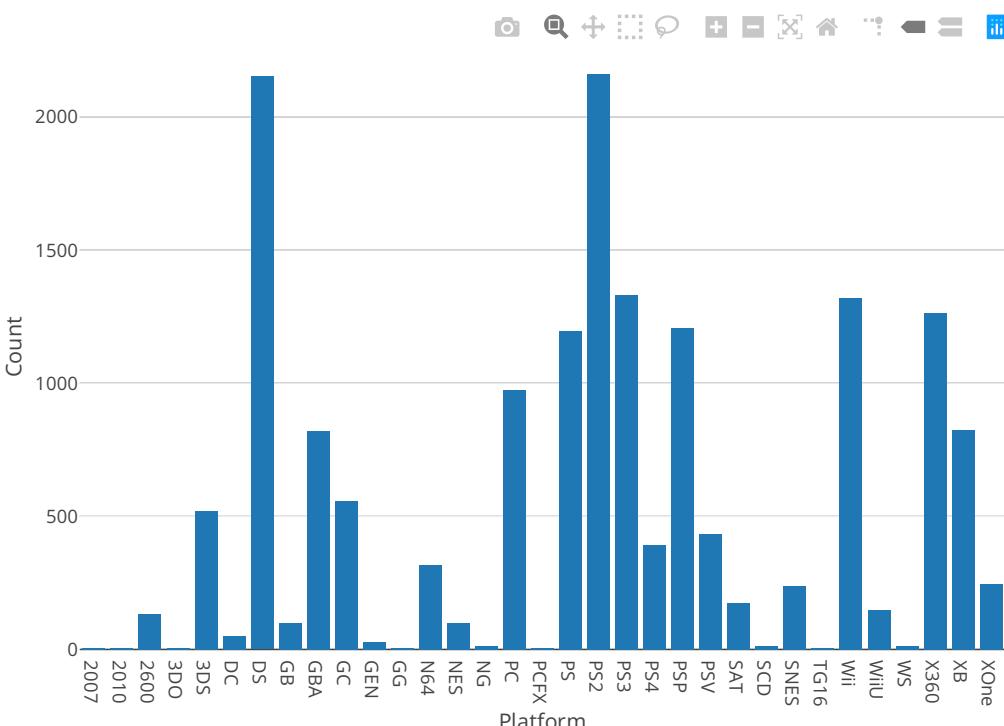


>>>> f20ead8b3f4b086b8313b0c724be842f4668860c

#From the bar chart above, Action Genre games yielded the highest sales globally. ##DC and PS were the platforms that benefitted with the highest sales at the global level. Although X360 and WiiU also strived at the global the global sales.

```
VIDEOGS%>%
group_by(Platform) %>%
summarise(Count = n()) %>%
plot_ly(x = ~Platform,
        y = ~Count,
        type = "bar")
```

<<<<< HEAD

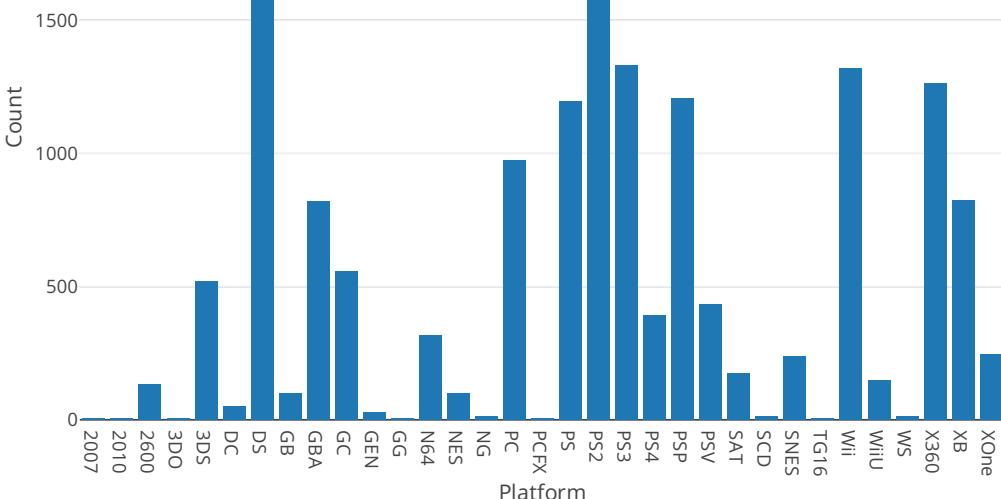


=====



2000

2000

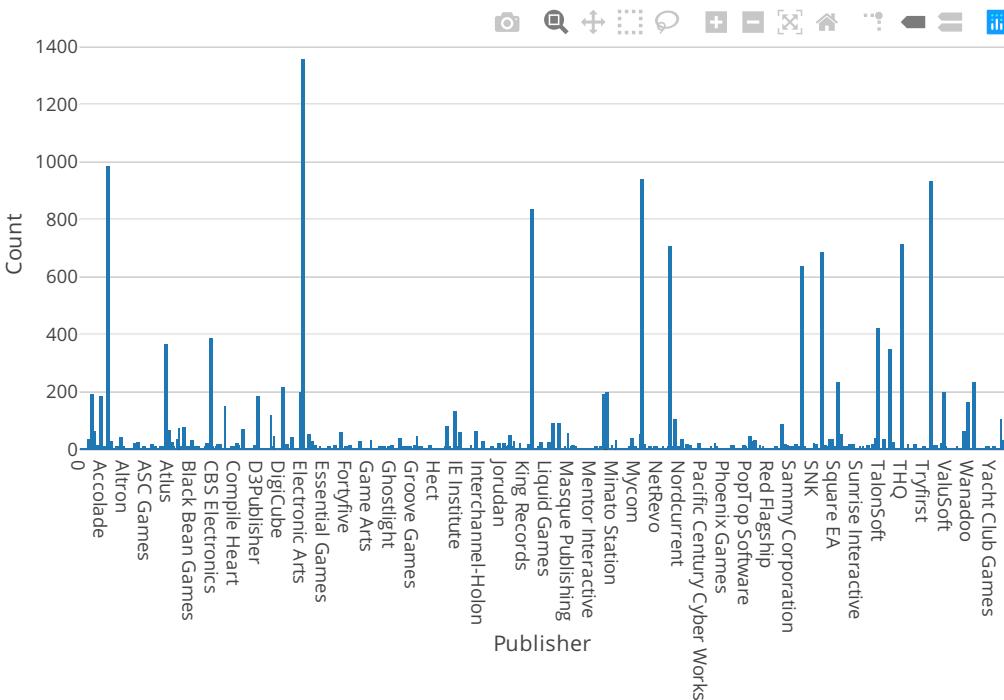


>>>>> f20ead8b3f4b086b8313b0c724be842f4668860c

#Sales by publishers #The bar chart revealed that Electronic Arts was the publisher with the highest sales.

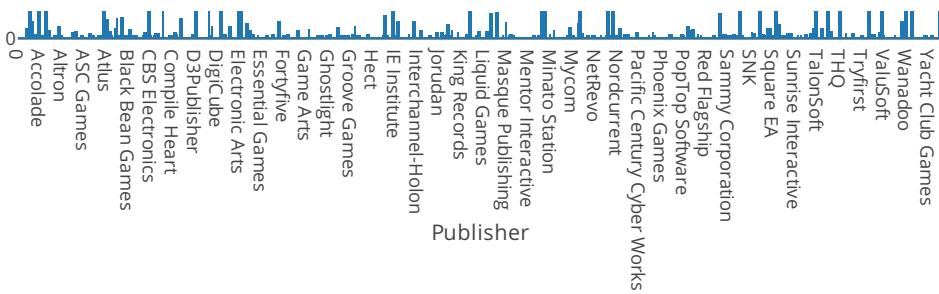
```
VIDEOGS%>%
  group_by(Publisher) %>%
  summarise(Count = n()) %>%
  plot_ly(x = ~Publisher,
          y = ~Count,
          type = "bar")
```

<<<<< HEAD



=====





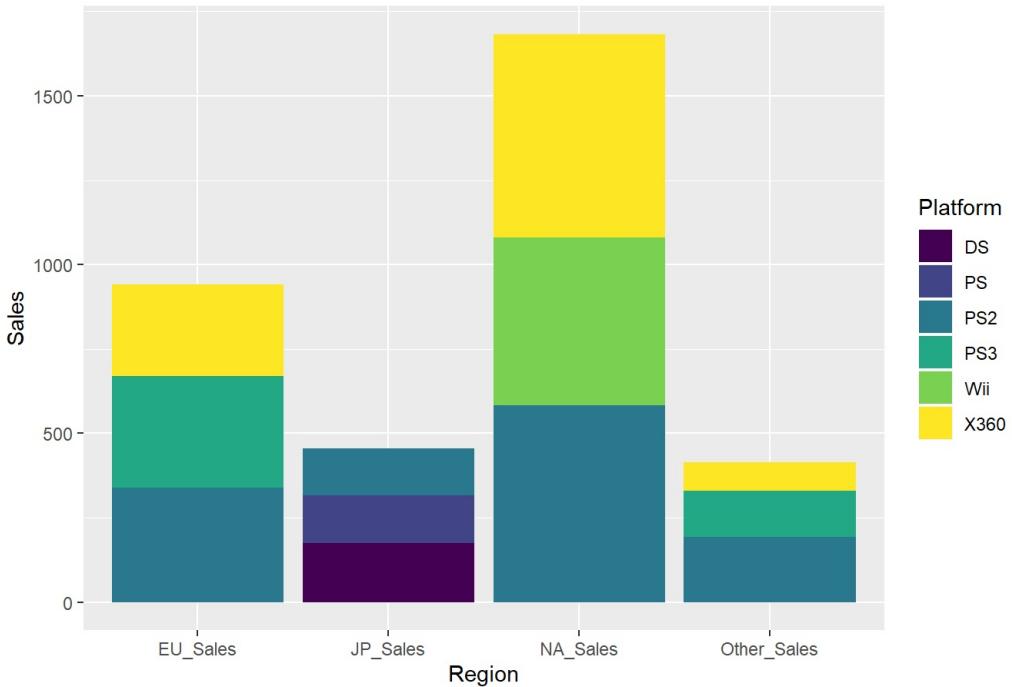
>>>>> f20ead8b3f4b086b8313b0c724be842f4668860c

#top sales by Platforms by region #The top platforms by sales in different region were DS, PS, PS2, PS3, WiiU and X360

```
VIDEOGS %>%
gather("Region", "Value", c("NA_Sales", "EU_Sales", "JP_Sales", "Other_Sales")) %>%
group_by(Region, Platform) %>%
summarize(Sales = sum(Value)) %>%
top_n(n = 3) %>%
ggplot(aes(x = Region, y = Sales, group = Region, fill = Platform)) +
geom_col(position = "stack") +
scale_fill_viridis(discrete = TRUE) +
labs(title = "Top Genre by Sales per Region")
```

```
## Selecting by Sales
```

Top Genre by Sales per Region

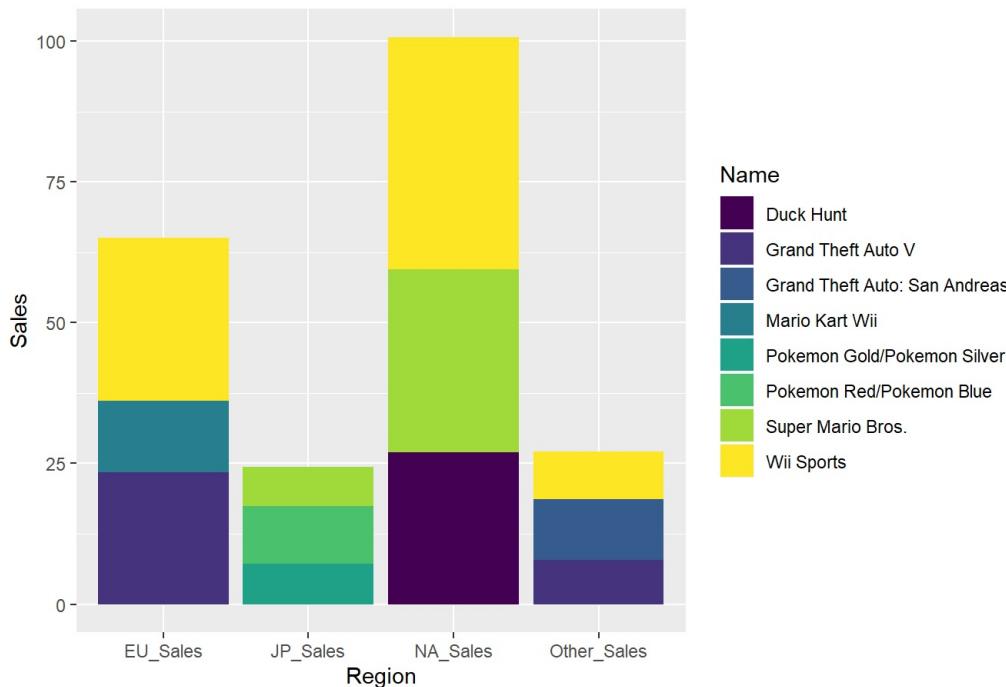


#Top sold games in the regions of the world

```
VIDEOGS %>%
gather("Region", "Value", c("NA_Sales", "EU_Sales", "JP_Sales", "Other_Sales")) %>%
group_by(Region, Name) %>%
summarize(Sales = sum(Value)) %>%
top_n(n = 3) %>%
ggplot(aes(x = Region, y = Sales, group = Region, fill = Name)) +
geom_col(position = "stack") +
scale_fill_viridis(discrete = TRUE) +
labs(title = "Top Genre by Sales per Region")
```

```
## Selecting by Sales
```

Top Genre by Sales per Region

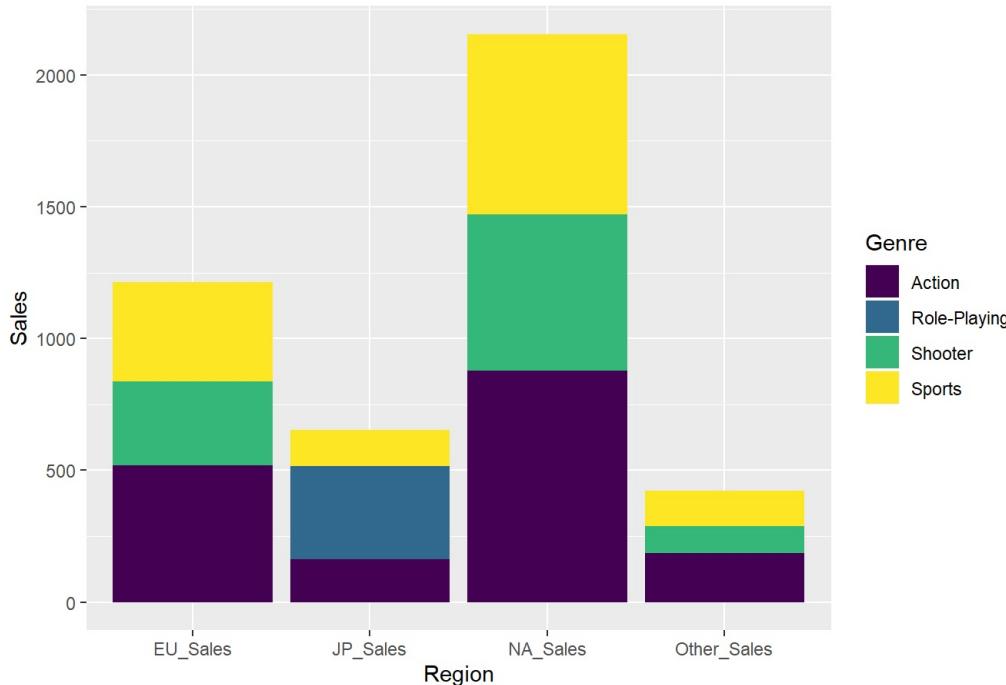


#Top sold genres according to regions #Action, Role-playing, Shooter and Sport

```
VIDEOGS %>%
gather("Region", "Value", c("NA_Sales", "EU_Sales", "JP_Sales", "Other_Sales")) %>%
group_by(Region, Genre) %>%
summarize(Sales = sum(Value)) %>%
top_n(n = 3) %>%
ggplot(aes(x = Region, y = Sales, group = Region, fill = Genre)) +
geom_col(position = "stack") +
scale_fill_viridis(discrete = TRUE) +
labs(title = "Top Genre by Sales per Region")
```

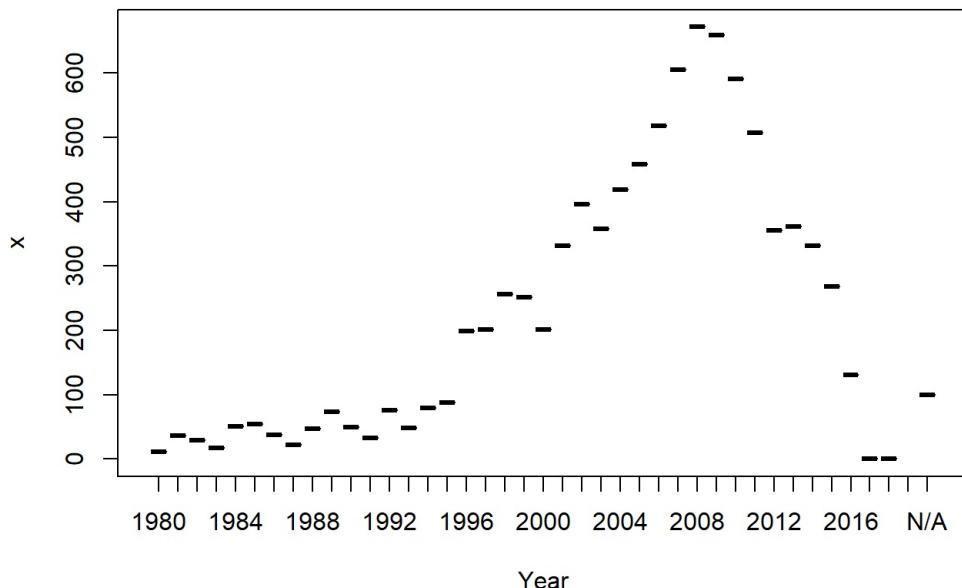
Selecting by Sales

Top Genre by Sales per Region



#total sales per year at global level

```
tot_year <- aggregate(VIDEOGS$Global_Sales, by=list(Year=VIDEOGS$Year), sum)
plot(tot_year)
```



```
data_frame(tot_year)
```

```
## Warning: `data_frame()` is deprecated, use `tibble()``.
## This warning is displayed once per session.
```

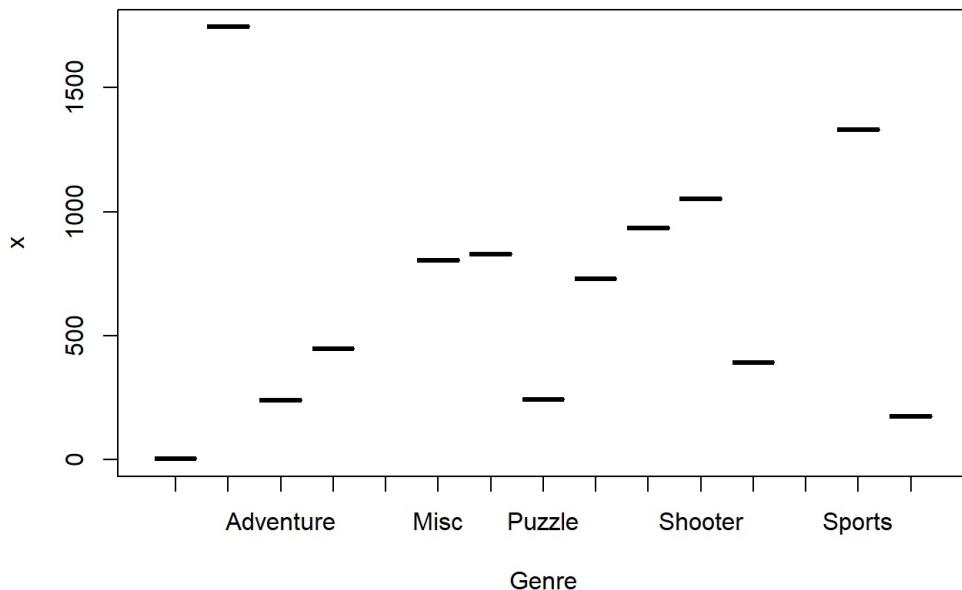
```
## # A tibble: 41 x 1
##   tot_year$Year     $x
##   <fct>      <dbl>
## 1 1980        11.4
## 2 1981        35.8
## 3 1982        28.9
## 4 1983        16.8
## 5 1984        50.4
## 6 1985        53.9
## 7 1986        37.1
## 8 1987        21.7
## 9 1988        47.2
## 10 1989       73.4
## # ... with 31 more rows
```

#Global Sales per genre

```
Glb_sales <- aggregate(VIDEOGS$Global_Sales, by=list(Genre=VIDEOGS$Genre), sum)
data.table(Glb_sales)
```

| | Genre | x |
|------------------------------------|--------------|---------|
| ## 1: | | 2.42 |
| ## 2: | Action | 1745.27 |
| ## 3: | Adventure | 237.57 |
| ## 4: | Fighting | 447.48 |
| ## 5: | Idea Factory | NA |
| ## 6: | Misc | 803.18 |
| ## 7: | Platform | 828.08 |
| ## 8: | Puzzle | 243.02 |
| ## 9: | Racing | 728.90 |
| ## 10: | Role-Playing | 934.40 |
| ## 11: | Shooter | 1052.94 |
| ## 12: | Simulation | 390.42 |
| ## 13: Sony Computer Entertainment | NA | |
| ## 14: | Sports | 1332.00 |
| ## 15: | Strategy | 174.50 |

```
plot(Glb_sales)
```



#number of games released per #the output revealed that the highest number of games was produced in 2008 with a total of 1427 games which was followed by the year 2007 with 1426.

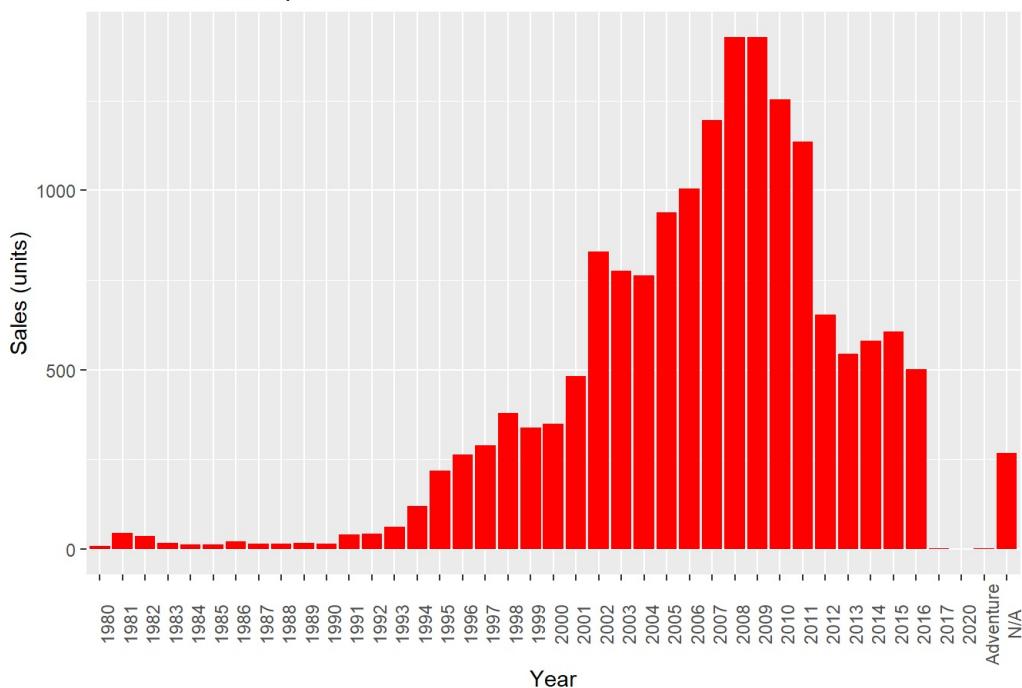
```
VIDEOGS %>%
  group_by(Year_of_Release) %>%
  summarize(Number_of_games_each_year = n())
```

```
## # A tibble: 41 x 2
##   Year_of_Release Number_of_games_each_year
##   <fct>                <int>
## 1 1980                  9
## 2 1981                 46
## 3 1982                 36
## 4 1983                 17
## 5 1984                 14
## 6 1985                 14
## 7 1986                 21
## 8 1987                 16
## 9 1988                 15
## 10 1989                17
## # ... with 31 more rows
```

#PLOT

```
VIDEOGS %>%
  group_by(Year_of_Release) %>%
  summarize(Number_of_games_each_year = n()) %>%
  ggplot(aes(x = Year_of_Release, y = Number_of_games_each_year)) +
  geom_col(fill = "red") +
  theme(axis.text.x = element_text(angle = 90)) +
  labs(title = "Games released per Year", x = "Year", y = "Sales (units)")
```

Games released per Year

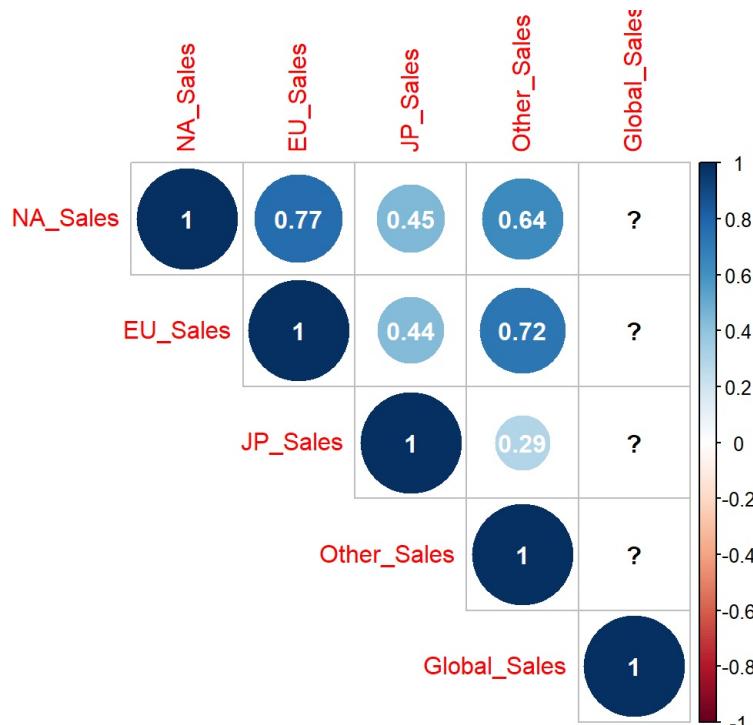


#a simple correlation matrix was carried out between sales

```
library(corrplot)

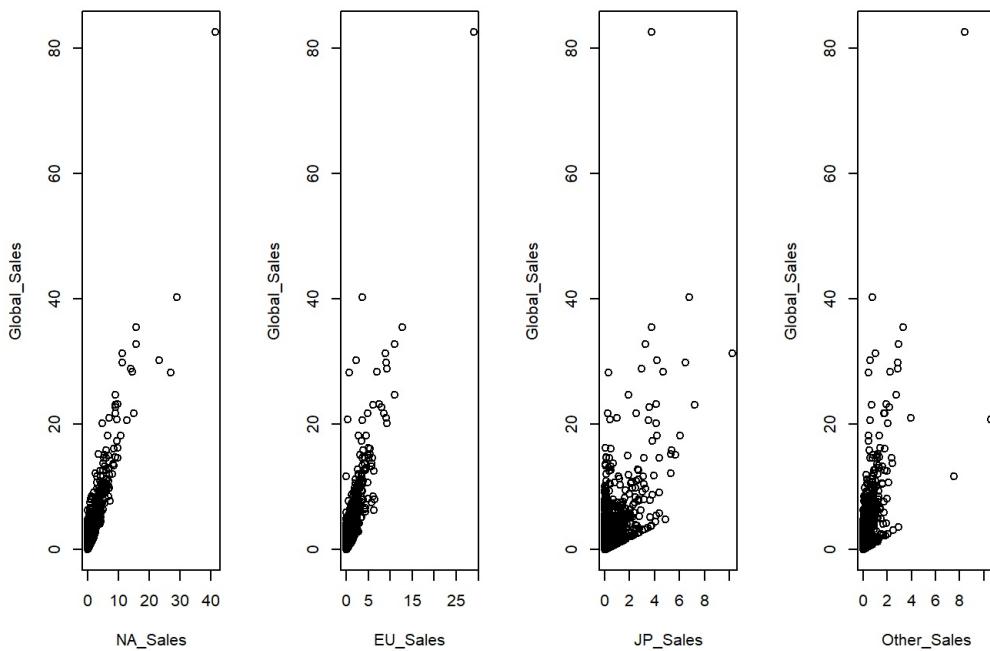
## corrplot 0.84 loaded

VIDEOGS[, c("NA_Sales", "EU_Sales", "JP_Sales", "Other_Sales", "Global_Sales")] %>%
  cor(method = "pearson") %>%
  corrplot::corrplot(addCoef.col = "white", type="upper")
```



#We considered the relationship between Sales in other regions against the global sales.

```
par(mfrow=c(1,4))
with(VIDEOGS, plot(NA_Sales, Global_Sales))
with(VIDEOGS, plot(EU_Sales, Global_Sales))
with(VIDEOGS, plot(JP_Sales, Global_Sales))
with(VIDEOGS, plot(Other_Sales, Global_Sales))
```



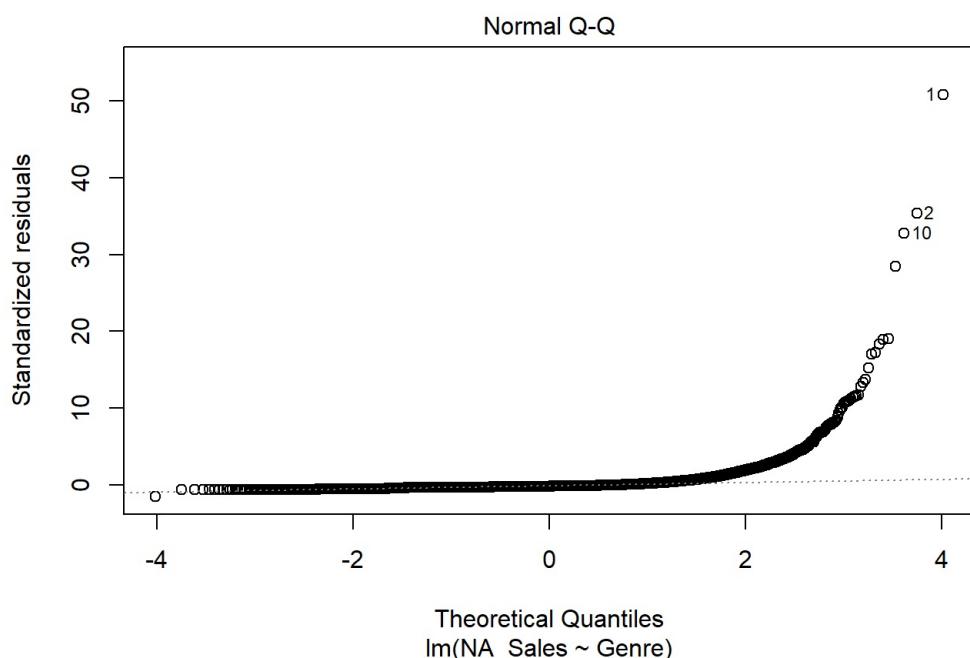
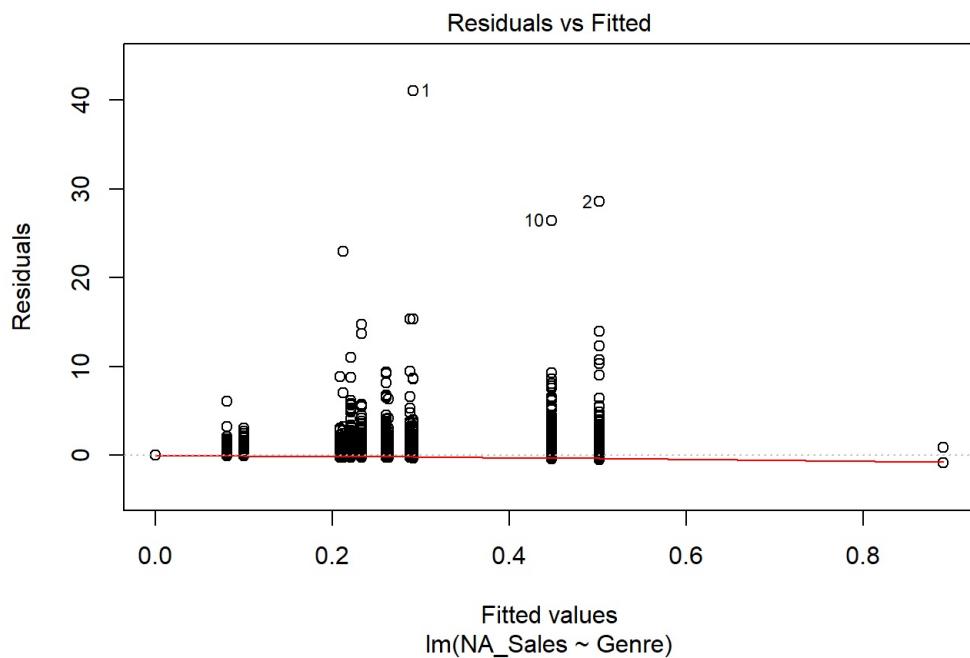
#The result revealed that there is linear relationship between regional sales and global sales. #we further considered linear models between sales in all the regions and genres in the study since action Genres have dominated sales. #North America

```
fit <- lm( NA_Sales ~ Genre, VIDEOGS)
summary(fit)
```

```
##
## Call:
## lm(formula = NA_Sales ~ Genre, data = VIDEOGS)
##
## Residuals:
##    Min     1Q Median     3Q    Max 
## -0.890 -0.233 -0.151 -0.013 41.069 
## 
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept) 0.8900    0.5711   1.558   0.119    
## GenreAction -0.6292    0.5712  -1.101   0.271    
## GenreAdventure -0.8091    0.5715  -1.416   0.157    
## GenreFighting -0.6269    0.5717  -1.096   0.273    
## GenreIdea Factory -0.8900    0.9891  -0.900   0.368    
## GenreMisc -0.6573    0.5714  -1.150   0.250    
## GenrePlatform -0.3883    0.5717  -0.679   0.497    
## GenrePuzzle -0.6782    0.5721  -1.185   0.236    
## GenreRacing -0.6023    0.5715  -1.054   0.292    
## GenreRole-Playing -0.6695    0.5715  -1.172   0.241    
## GenreShooter -0.4424    0.5715  -0.774   0.439    
## GenreSimulation -0.6815    0.5717  -1.192   0.233    
## GenreSony Computer Entertainment -0.8900    0.9891  -0.900   0.368    
## GenreSports -0.5985    0.5713  -1.048   0.295    
## GenreStrategy -0.7896    0.5719  -1.381   0.167    
## 
## Residual standard error: 0.8076 on 16704 degrees of freedom
## Multiple R-squared:  0.01527,   Adjusted R-squared:  0.01444 
## F-statistic:  18.5 on 14 and 16704 DF,  p-value: < 2.2e-16
```

```
plot(fit)
```

```
## Warning: not plotting observations with leverage one:
##      13587
```



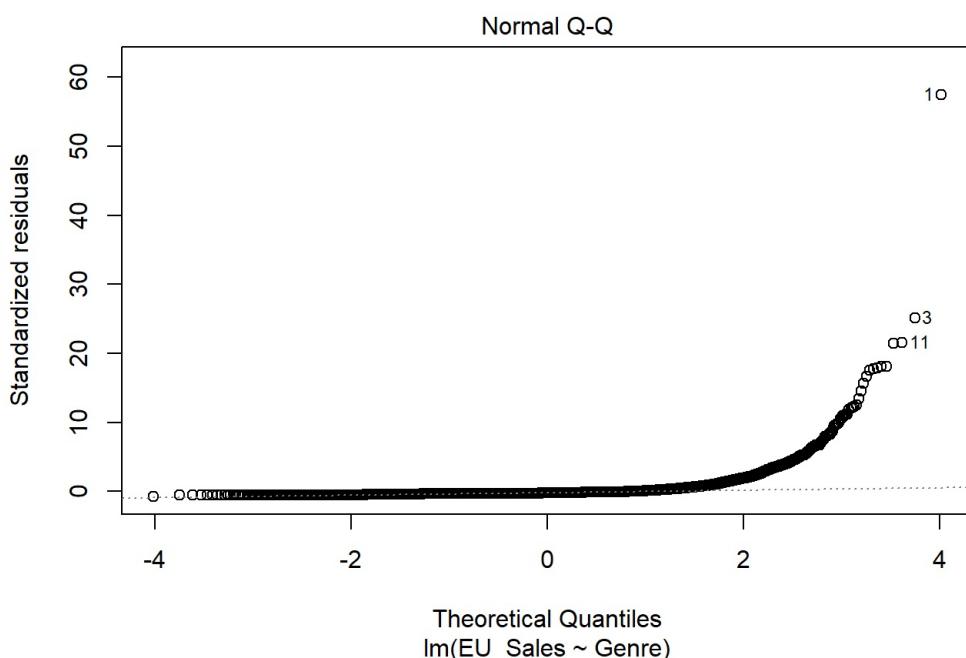
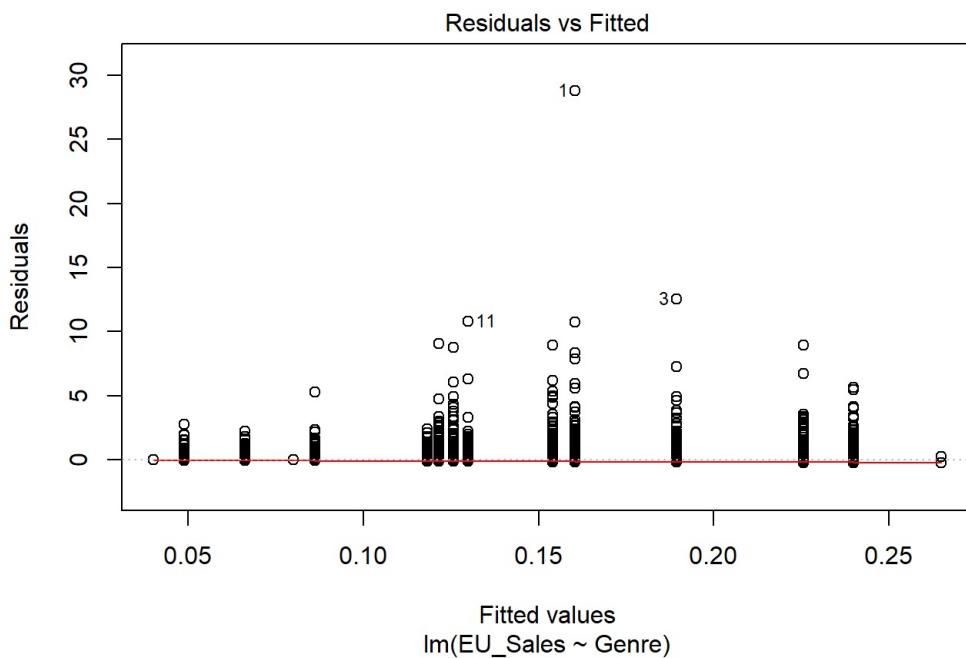
```
## Warning: not plotting observations with leverage one:  
##      13587
```



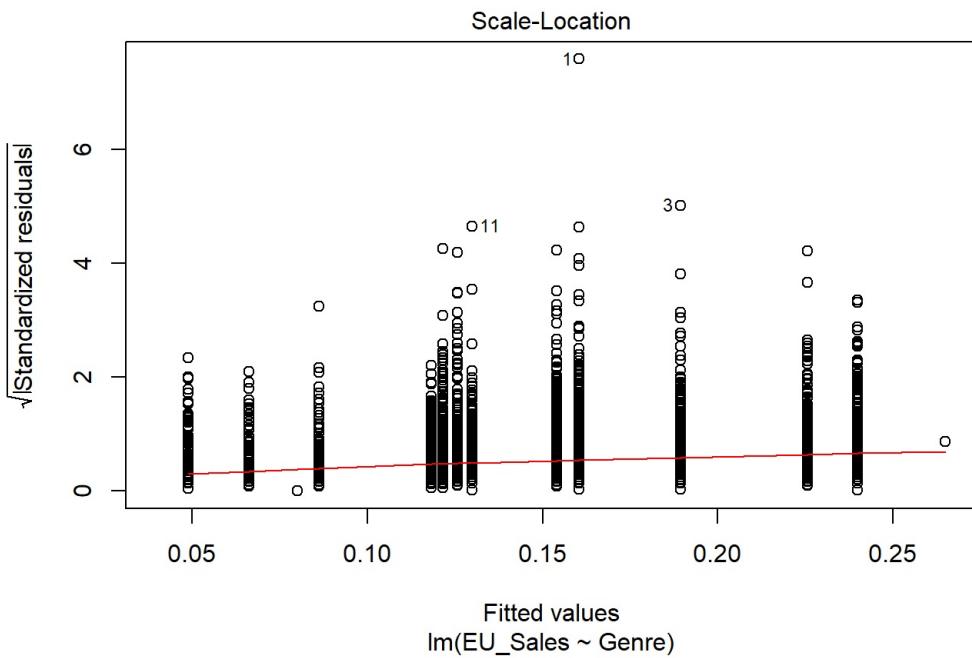
```
##  
## Call:  
## lm(formula = EU_Sales ~ Genre, data = VIDEOGS)  
##  
## Residuals:  
##      Min      1Q  Median      3Q     Max  
## -0.2650 -0.1440 -0.1058 -0.0294 28.7995  
##  
## Coefficients:  
##                               Estimate Std. Error t value Pr(>|t|)  
## (Intercept)                 0.26500  0.35427   0.748   0.454  
## GenreAction                -0.11096  0.35437  -0.313   0.754  
## GenreAdventure              -0.21616  0.35454  -0.610   0.542  
## GenreFighting               -0.14683  0.35469  -0.414   0.679  
## GenreIdea Factory           -0.22500  0.61361  -0.367   0.714  
## GenreMisc                  -0.14343  0.35447  -0.405   0.686  
## GenrePlatform               -0.03938  0.35467  -0.111   0.912  
## GenrePuzzle                -0.17878  0.35488  -0.504   0.614  
## GenreRacing                -0.07564  0.35455  -0.213   0.831  
## GenreRole-Playing           -0.13919  0.35451  -0.393   0.695  
## GenreShooter                -0.02514  0.35454  -0.071   0.943  
## GenreSimulation              -0.13511  0.35467  -0.381   0.703  
## GenreSony Computer Entertainment -0.18500  0.61361  -0.301   0.763  
## GenreSports                 -0.10453  0.35442  -0.295   0.768  
## GenreStrategy                -0.19887  0.35479  -0.561   0.575  
##  
## Residual standard error: 0.501 on 16704 degrees of freedom  
## Multiple R-squared:  0.009829,   Adjusted R-squared:  0.009  
## F-statistic: 11.84 on 14 and 16704 DF,  p-value: < 2.2e-16
```

```
plot(fit)
```

```
## Warning: not plotting observations with leverage one:  
##       13587
```

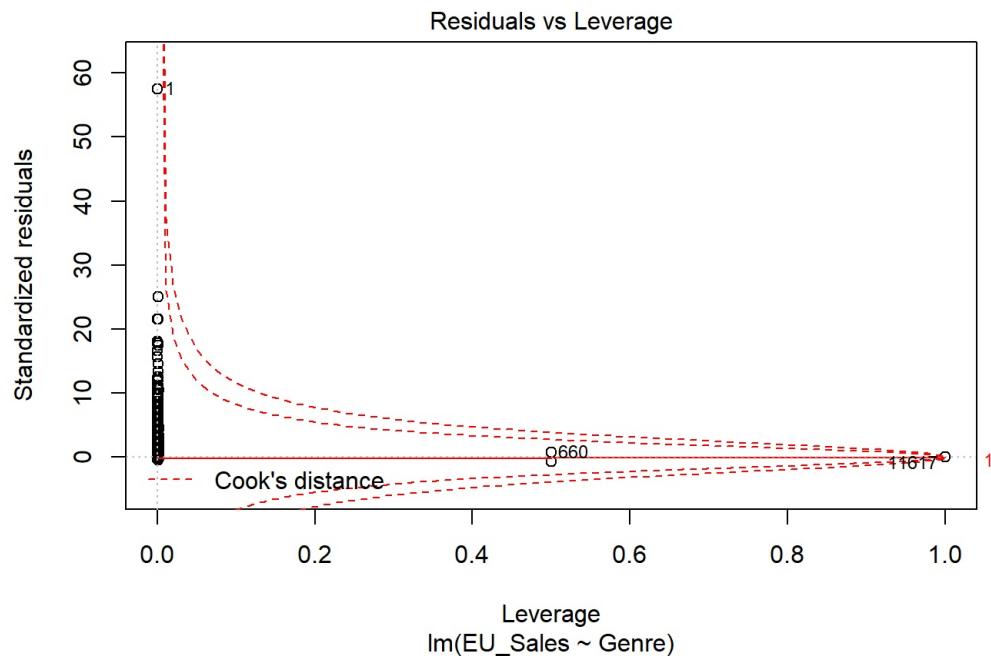


```
## Warning: not plotting observations with leverage one:  
##      13587
```



```
## Warning in sqrt(crit * p * (1 - hh)/hh): 产生了NaNs
```

```
## Warning in sqrt(crit * p * (1 - hh)/hh): 产生了NaNs
```



```
#linear relationship between sales in Japan and Genre
```

```
fit <- lm( JP_Sales ~ Genre, VIDEOGS)
summary(fit)
```

```

## 
## Call:
## lm(formula = JP_Sales ~ Genre, data = VIDEOGS)
##
## Residuals:
##    Min      1Q  Median      3Q     Max
## -0.2370 -0.0618 -0.0479 -0.0279  9.9830
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)    
## (Intercept)                0.01500   0.21474   0.070   0.944    
## GenreAction                 0.03291   0.21480   0.153   0.878    
## GenreAdventure              0.02511   0.21491   0.117   0.907    
## GenreFighting                0.08804   0.21499   0.409   0.682    
## GenreIdea Factory            -0.01500   0.37194  -0.040   0.968    
## GenreMisc                   0.04678   0.21486   0.218   0.828    
## GenrePlatform                0.13233   0.21498   0.616   0.538    
## GenrePuzzle                  0.08381   0.21511   0.390   0.697    
## GenreRacing                  0.03040   0.21491   0.141   0.887    
## GenreRole-Playing             0.22197   0.21488   1.033   0.302    
## GenreShooter                  0.01430   0.21490   0.067   0.947    
## GenreSimulation               0.05800   0.21499   0.270   0.787    
## GenreSony Computer Entertainment -0.01500   0.37194  -0.040   0.968    
## GenreSports                   0.04273   0.21483   0.199   0.842    
## GenreStrategy                  0.05771   0.21505   0.268   0.788    
## 
## Residual standard error: 0.3037 on 16704 degrees of freedom
## Multiple R-squared:  0.03376,    Adjusted R-squared:  0.03295 
## F-statistic: 41.69 on 14 and 16704 DF,  p-value: < 2.2e-16

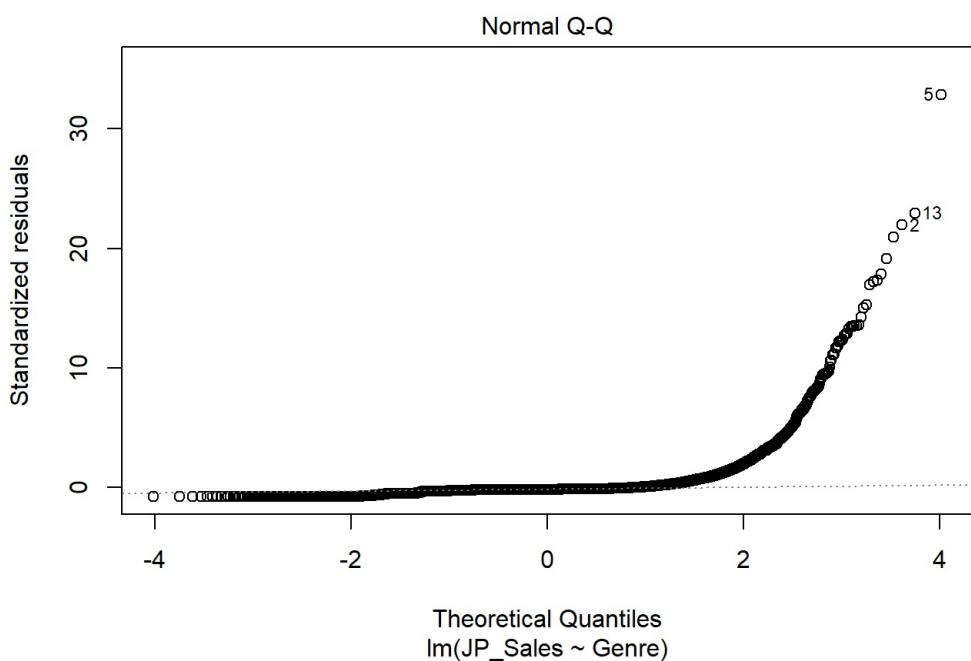
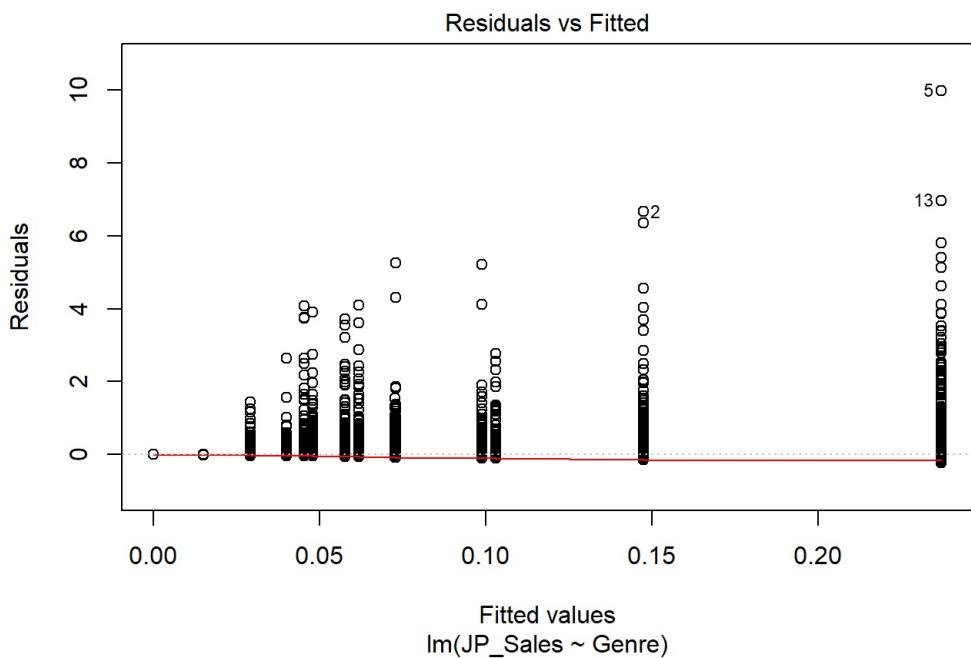
```

```
plot(fit)
```

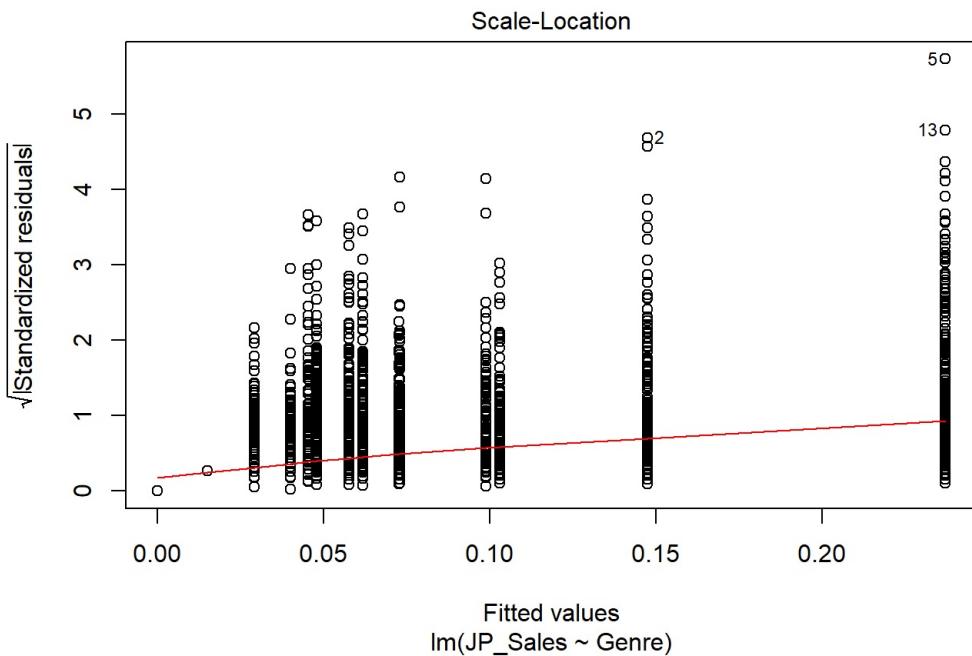
```

## Warning: not plotting observations with leverage one:
##       13587

```

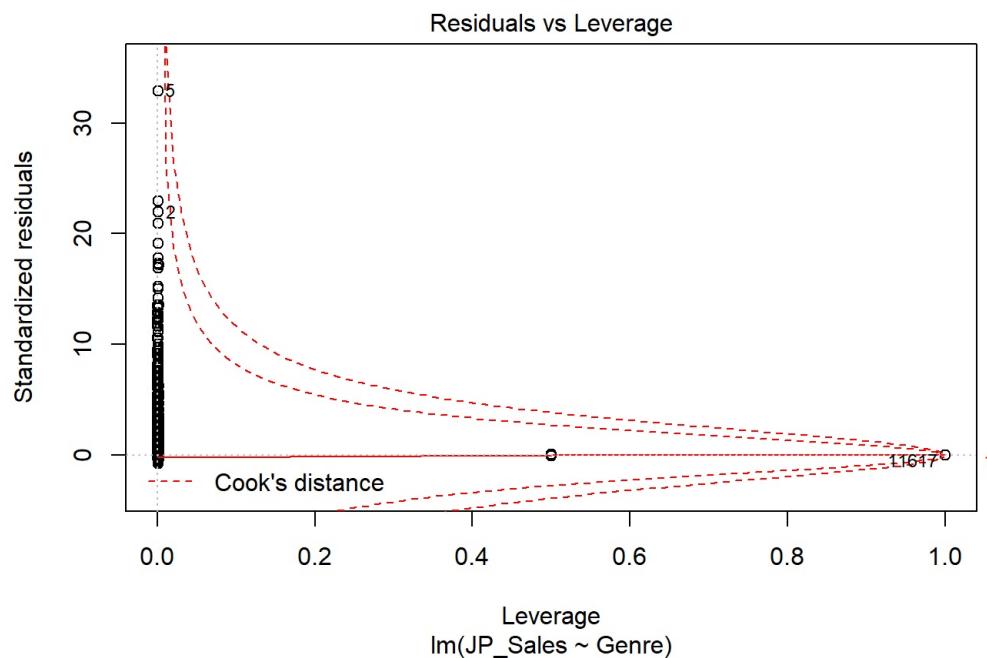


```
## Warning: not plotting observations with leverage one:  
##      13587
```



```
## Warning in sqrt(crit * p * (1 - hh)/hh): 产生了NaNs
```

```
## Warning in sqrt(crit * p * (1 - hh)/hh): 产生了NaNs
```



#linear relationship between other sales and Genre

```
fit <- lm(Other_Sales ~ Genre, VIDEOSGS)
summary(fit)
```

```

## 
## Call:
## lm(formula = Other_Sales ~ Genre, data = VIDEOGS)
##
## Residuals:
##    Min      1Q  Median      3Q     Max
## -0.0787 -0.0473 -0.0325 -0.0073 10.5152
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)    
## (Intercept)          4.000e-02  1.315e-01   0.304   0.761    
## GenreAction          1.478e-02  1.316e-01   0.112   0.911    
## GenreAdventure       -2.733e-02 1.316e-01  -0.208   0.836    
## GenreFighting        2.827e-03 1.317e-01   0.021   0.983    
## GenreIdea Factory    8.082e-12 2.278e-01   0.000   1.000    
## GenreMisc            2.509e-03 1.316e-01   0.019   0.985    
## GenrePlatform         1.753e-02 1.317e-01   0.133   0.894    
## GenrePuzzle          -1.866e-02 1.317e-01  -0.142   0.887    
## GenreRacing           2.093e-02 1.316e-01   0.159   0.874    
## GenreRole-Playing    -2.467e-04 1.316e-01  -0.002   0.999    
## GenreShooter          3.869e-02 1.316e-01   0.294   0.769    
## GenreSimulation       -4.817e-03 1.317e-01  -0.037   0.971    
## GenreSony Computer Entertainment 4.000e-02 2.278e-01   0.176   0.861    
## GenreSports            1.729e-02 1.316e-01   0.131   0.895    
## GenreStrategy          -2.411e-02 1.317e-01  -0.183   0.855    
## 
## Residual standard error: 0.186 on 16704 degrees of freedom
## Multiple R-squared:  0.00849,   Adjusted R-squared:  0.007659 
## F-statistic: 10.22 on 14 and 16704 DF,  p-value: < 2.2e-16

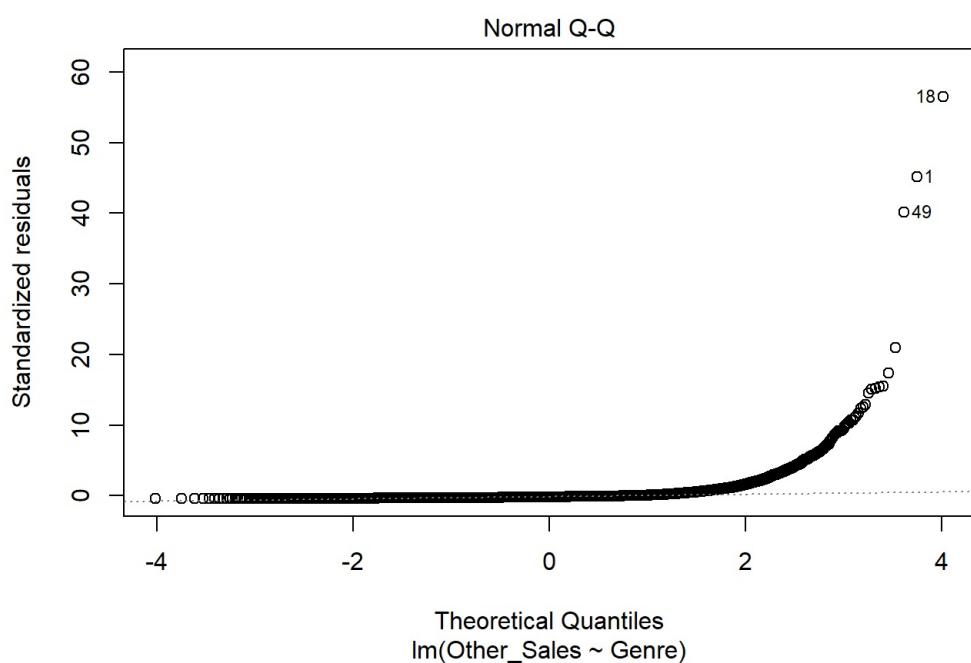
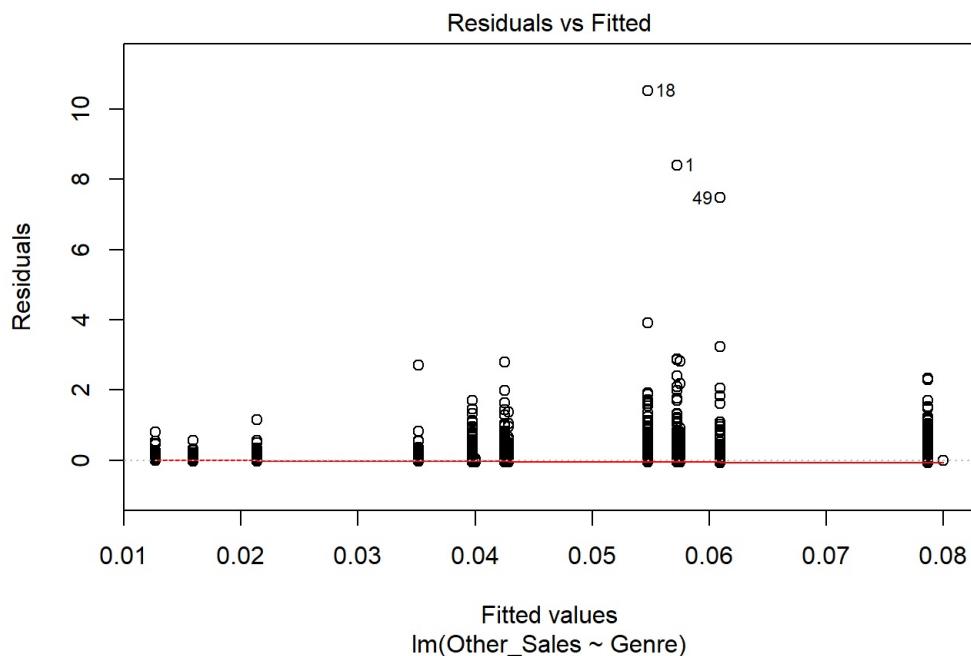
```

```
plot(fit)
```

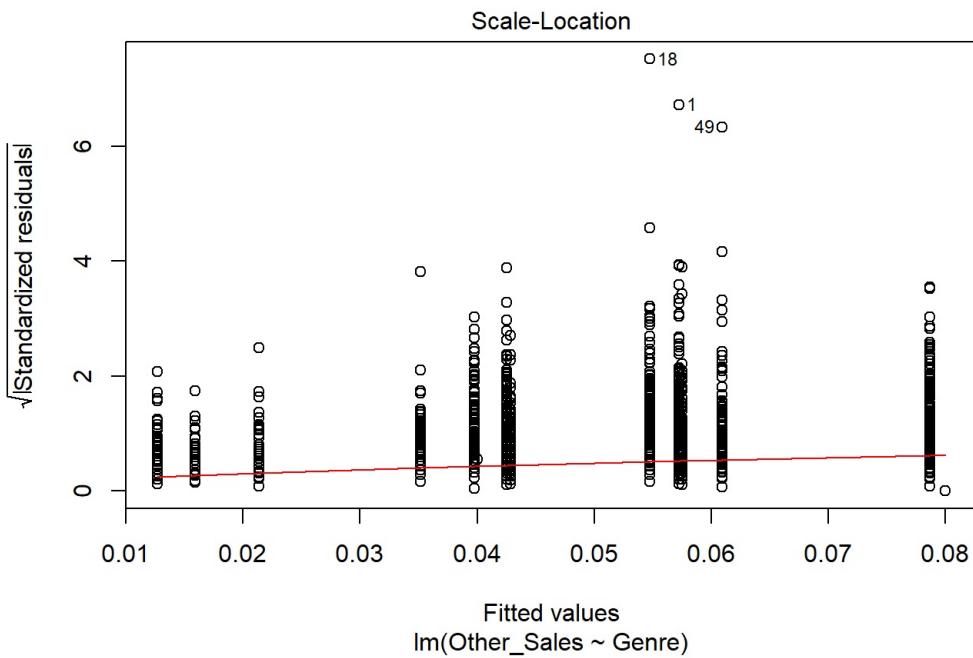
```

## Warning: not plotting observations with leverage one:
## 13587

```

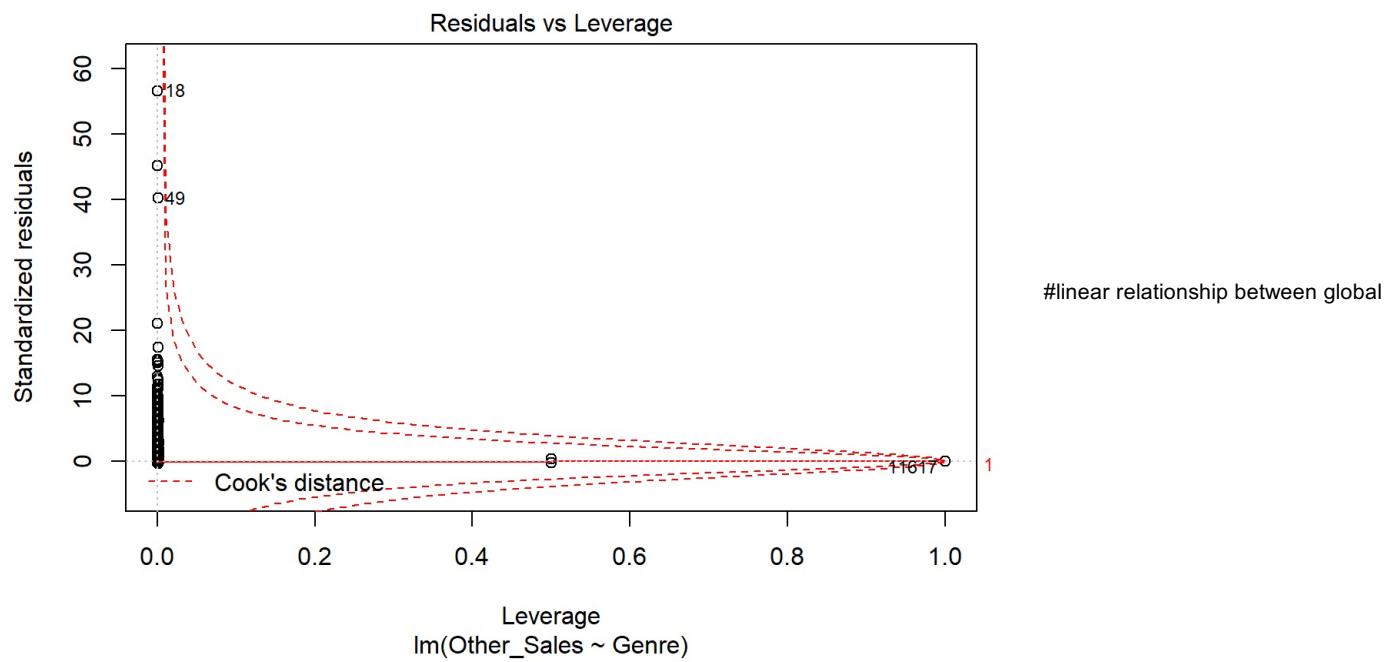


```
## Warning: not plotting observations with leverage one:  
##      13587
```



```
## Warning in sqrt(crit * p * (1 - hh)/hh): 产生了NaNs
```

```
## Warning in sqrt(crit * p * (1 - hh)/hh): 产生了NaNs
```

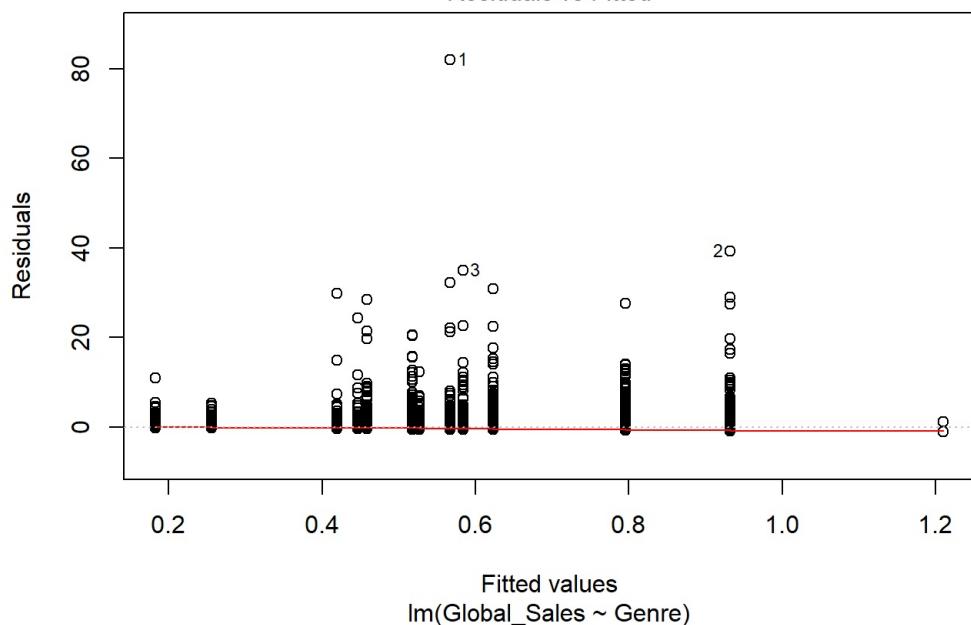


```
fit <- lm(Global_Sales ~ Genre, VIDEOS)
summary(fit)
```

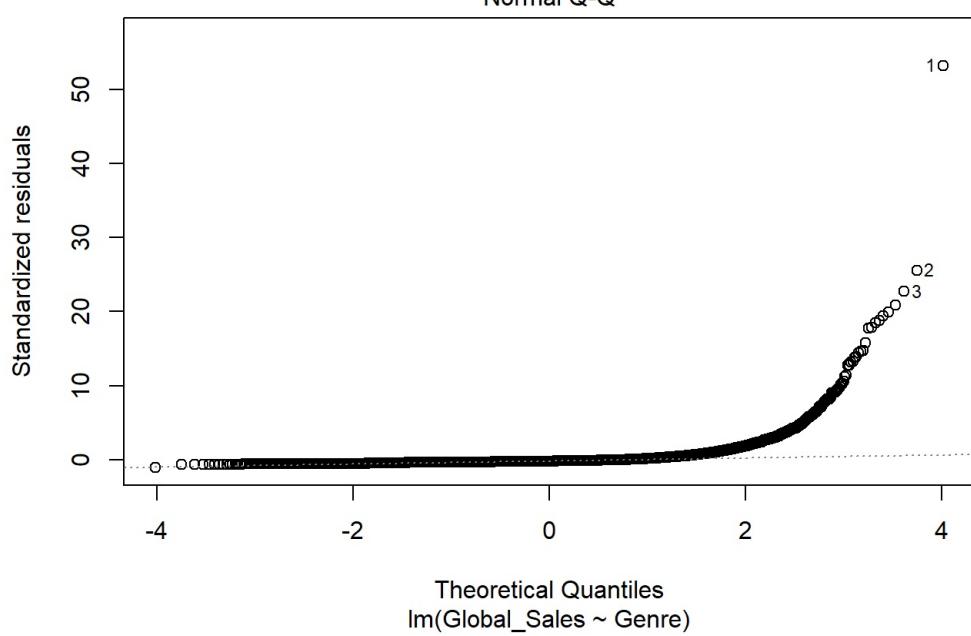
```
##  
## Call:  
## lm(formula = Global_Sales ~ Genre, data = VIDEOGS)  
##  
## Residuals:  
##    Min      1Q Median      3Q     Max  
## -1.180 -0.457 -0.307 -0.039 81.963  
##  
## Coefficients:  
##              Estimate Std. Error t value Pr(>|t|)  
## (Intercept) 1.2100    1.0883   1.112   0.266  
## GenreAction -0.6921    1.0886  -0.636   0.525  
## GenreAdventure -1.0274    1.0891  -0.943   0.346  
## GenreFighting -0.6829    1.0896  -0.627   0.531  
## GenreMisc -0.7510    1.0889  -0.690   0.490  
## GenrePlatform -0.2775    1.0895  -0.255   0.799  
## GenrePuzzle -0.7910    1.0902  -0.726   0.468  
## GenreRacing -0.6264    1.0892  -0.575   0.565  
## GenreRole-Playing -0.5871    1.0890  -0.539   0.590  
## GenreShooter -0.4141    1.0891  -0.380   0.704  
## GenreSimulation -0.7633    1.0895  -0.701   0.484  
## GenreSports -0.6427    1.0888  -0.590   0.555  
## GenreStrategy -0.9545    1.0899  -0.876   0.381  
##  
## Residual standard error: 1.539 on 16704 degrees of freedom  
## (2 observations deleted due to missingness)  
## Multiple R-squared:  0.01221,    Adjusted R-squared:  0.0115  
## F-statistic:  17.2 on 12 and 16704 DF,  p-value: < 2.2e-16
```

```
plot(fit)
```

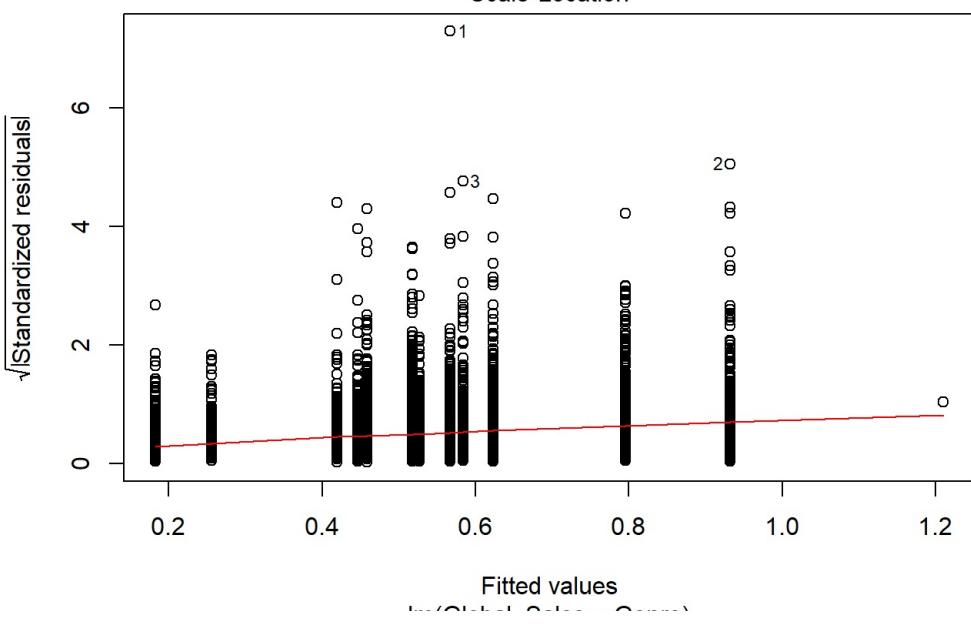
Residuals vs Fitted



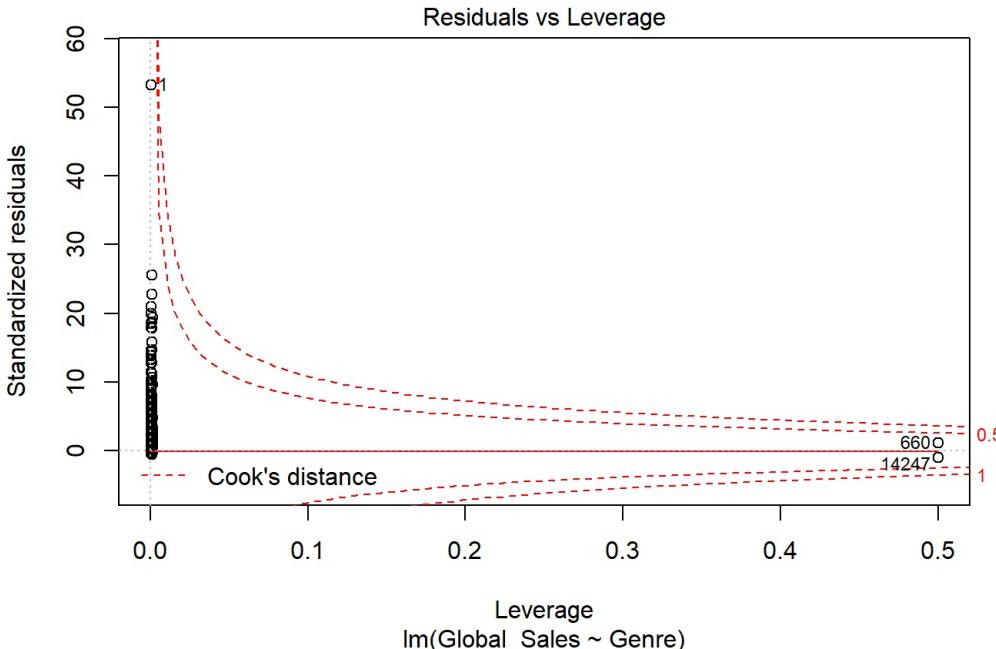
Normal Q-Q



Scale-Location



lm(Global_Sales ~ Genre)

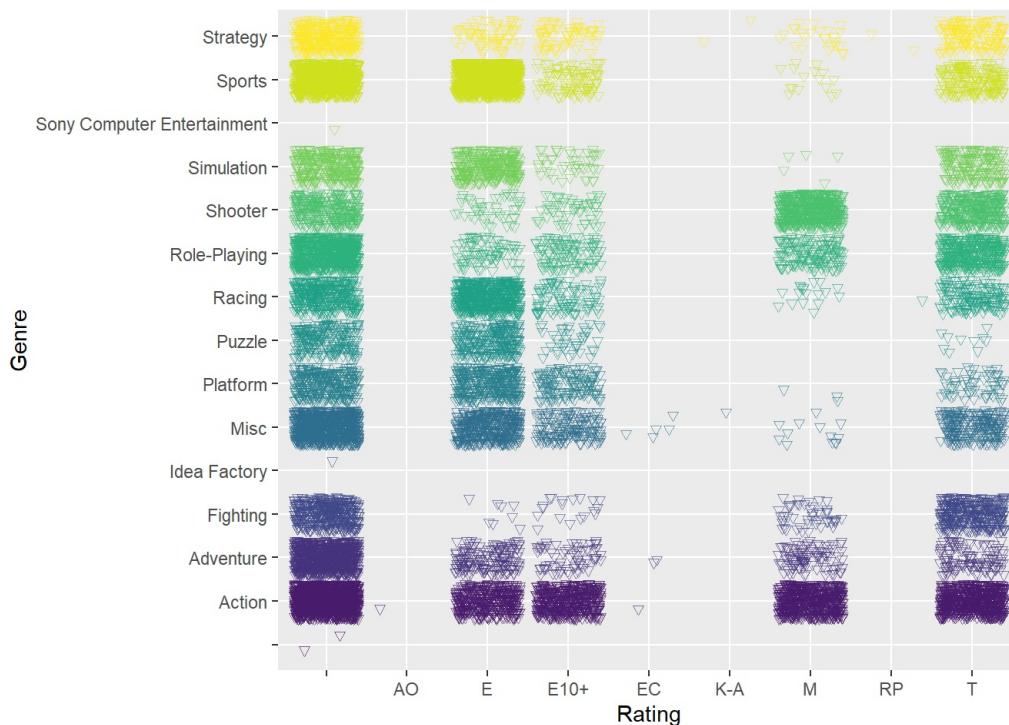


#with reference to the fitted plots, it

is obvious that genres has no influence on the sales of games in all the regions. #we then looked at the relationship between sales and Genre and Rating #The result revealed a strong relationship between sales and Genre.

```
VIDEOGS %>%
  ggplot(aes(x = Rating, y = Genre, col = Genre)) +
  geom_jitter(alpha = 0.6, pch = 25) +
  theme(legend.position = "none") +
  scale_color_viridis(discrete = TRUE)
```

<<<<< HEAD



DISCUSSION The main focus of the

project is to evaluate video game sales and see how changes have occurred over the decades in the industry. The result revealed the following, 1. In all the regions of the world, the most popular genres according to sales are Action, sports and Shooting games. 2. There are linear relationship between sales and Genres. 3. There is a relationship between North America region sales and European Union region sales, Other sales and Global sales with pearson correlation values shown in the plot below. That is to say that any change in the sales of games by either genre, developer or publisher at regional level contributes tremendously to the global sales of any of the variables.

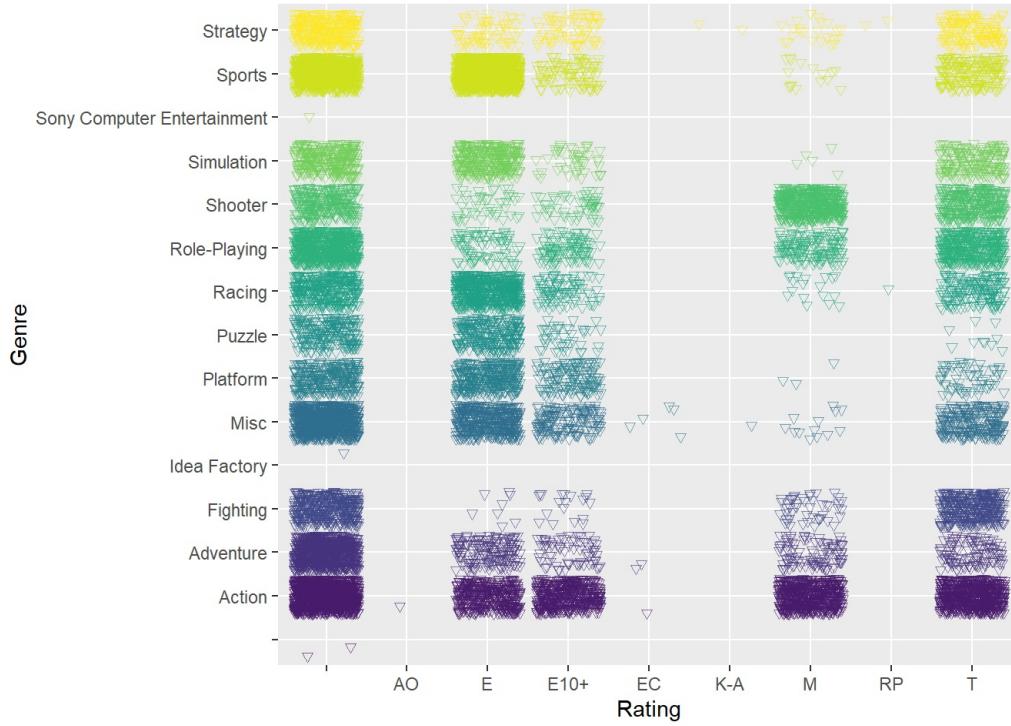
CONCLUSION In conclusion, the study revealed that there is a relationship between sales. This means that increase or decrease in regional sales will result to proportional increase in global sales. The study further revealed that, a. the popularity of Action, Sports and Shooter genres has shown tremendous growth over 20 years. b. DS, PS and XBOX and Wii publishers were among the top developers that have strived over the years in the industries. c. North America was observed as the region with the highest number of sales followed by European Union region. d. There was no change in sales with genre in the regions. e. The highest number of sales in past 30 years was obtained in year 2008 f. The highest number of games produced was in 2008 with a total of 1427 games and was closely followed by the year 2007 with 1426. g. and video game industries have grown exponentially over the years with top publisher still in remaining in the business while the weak ones have fizzled out

of the industry. RECOMMENDATIONS 1. The striving platforms should consider other regions with affordable Action genres. The game industry will keep growing in demand with the growing population of the world and such the demand of games will grow proportional with this. To this i recommend that more developers should take advantage of the market in the nearest future.

REFERENCES 1. https://www.kaggle.com/rush4ratio/video-game-sales-with-ratings/data#Video_Games_Sales_as_at_22_Dec_2016.csv
[\(https://www.kaggle.com/rush4ratio/video-game-sales-with-ratings/data#Video_Games_Sales_as_at_22_Dec_2016.csv\)](https://www.kaggle.com/rush4ratio/video-game-sales-with-ratings/data#Video_Games_Sales_as_at_22_Dec_2016.csv)

2. <https://www.kaggle.com/umeshnarayanappa/explore-video-games-sales> (<https://www.kaggle.com/umeshnarayanappa/explore-video-games-sales>)
http://scholarship.claremont.edu/cgi/viewcontent.cgi?article=1972&context=cmc_theses
http://scholarship.claremont.edu/cgi/viewcontent.cgi?article=1972&context=cmc_theses
3. <https://www.kaggle.com/umeshnarayanappa/explore-video-games-sales> (<https://www.kaggle.com/umeshnarayanappa/explore-video-games-sales>)
4. https://rstudio-pubs-static.s3.amazonaws.com/346100_d6f3f54c8f454f918456dea6b23ce7b0.html (https://rstudio-pubs-static.s3.amazonaws.com/346100_d6f3f54c8f454f918456dea6b23ce7b0.html)

=====



>>>>> f20ead8b3f4b086b8313b0c724be842f4668860c