Discover and Data Preparation

Aaron Nicanor's Data Science Portfolio Project

Introduction

My decision to work on this dataset was because I found it to be far more interesting and manageable than the topic I had chosen prior. Originally, I had wanted to analyse employment data, but I began to find other datasets that spoke to me far more than the ones I had originally been aiming for.

With that said, I will be analysing video game sales with relation to their ratings and reviews. I find this to be a compelling topic because I'd like to see what types of games would garner larger sales and to see how perceived quality of said games would affect their marketability.

Source Quality

This dataset I'm using can be found at https://www.kaggle.com/rush4ratio/video-game-sales-with-ratings/. This acts as an extension to another dataset published which can be found at https://www.kaggle.com/gregorut/videogamesales. In the original dataset, it aimed to represent games with sales that sold greater than 100,000 copies. This was generated by a scrape of vgchartz.com which is a site that tracks weekly sales figures of console hardware. In the current dataset I'll be using, additional data was appended to the prior set by adding metacritic scores to the games listed. Metacritic is a site compiling reviews from critics and users alike.

Possible faults that can come from using this data is that vgchartz has fallen under some criticism for the validity of how they obtain and verify their data. As a result, some sales figures I receive from my dataset may not be completely accurate. In addition, not all the games listed on this dataset have metacritic scores attached to them, so I may have some variables that will be missing review scores. Despite that, there are still about 6,900 complete cases.

Dataset

To begin, we downloaded the dataset into my project folder and loaded it in as a dataset named "InitialData". We also load in the libraries for tidyverse and ggplot2 since we'll need these in the future.

```
suppressMessages(library("tidyverse"))
suppressMessages(library("dplyr"))
suppressMessages(library("tidyr"))
suppressMessages(library("ggplot2"))
InitialData <- read.csv("Video_Games_Sales_as_at_22_Dec_2016.csv", header=TRUE, sep=",")
```

Column Variables

- **Title**: The name of the game to aid in identification.
- **Platform**: Lists the console platform in which the game was released on.
- **Year of Release**: The year the game came out.
- **Genre**: The genre that the game belongs to. Aids in categorizing.
- **Publisher**: The company that published the game. Publisher notariety may aid in game sales.
- **NA Sales**: Sales figures within North America in millions.
- **EU Sales**: Sales figures within Europe in millions.
- **JP Sales**: Sales figures within Japan in millions.
- **Other Sales**: Sales figures for the rest of the world in millions.
- **Global Sales**: Total worldwide sales in millions.
- **Critic Score**: The aggregate critic ratings compiled by the metacritic staff.
- **Critic Count**: The number of critics used to calculate the critic review rating
- **User Score**: The aggregate rating given to a game by Metacritic's subscribers.
- **User Count**: The number of users who gave their review rating on the game.
- **Developer**: The Party responsible for creating the game. Developer notoriety may aid in a game's sales.
- **Rating**: The ESRB rating. This could factor into a game's audience appeal.

Data Organization

To begin, we needed to organize and clean up our data.

```
#We'll first need to rename the column names within the dataset to match the variables names we had listed a
bove. This is mostly to aid in readability.
names(InitialData) <- gsub(x = names(InitialData), pattern = "\\_", replacement = " ")
colnames(InitialData)[colnames(InitialData)=="Name"] <- "Title"

#We only want to view data that is complete, so we remove any data that is missing a metacritic score.
InitialData <- InitialData[rowSums(is.na(InitialData[ , ])) == 0, ]

#We had to change the critic scoring so that it matched user scoring's out of 10
InitialData$`Critic Score` <- InitialData[,11]/10

#We need to specify which columns are factors for our dataset. These factors are our categorical data.
InitialData$Rating <- as.factor(InitialData$Rating)
InitialData$Genre <- as.factor(InitialData$Genre)
InitialData$Platform <- as.factor(InitialData$Platform)
InitialData$`User Score` <- as.factor(InitialData$`User Score`)
InitialData$`Critic Score` <- as.factor(InitialData$`Critic Score`)
```

Finally, we chose to separate our data into eight tables and subdivide them into two different categories: user based data, and critic based data. Users have a much larger pool of opinions that critics do, therefore making a distinction between the two would be helpful during analysis.

```
#We chose to differentiate the sales figures between the 3 largest markets for video games along with the gl
obal sales figures.

#North America
UserBasedDataNA <- tibble(Title=InitialData$Title,
                          Publisher=InitialData$Publisher,
                          Developer=InitialData$Developer,
                          `Year of Release`=InitialData$`Year of Release`,
                          Platform=InitialData$Platform,
                          Rating=InitialData$Rating,
                          Genre=InitialData$Genre,
                          `NA Sales`=InitialData$`NA Sales`,
                          `User Score`=InitialData$`User Score`,
                          `User Count`=InitialData$`User Count`)
CriticBasedDataNA <- tibble(Title=InitialData$Title,
                          Publisher=InitialData$Publisher,
                          Developer=InitialData$Developer,
                          `Year of Release`=InitialData$`Year of Release`,
                          Platform=InitialData$Platform,
                          Rating=InitialData$Rating,
                          Genre=InitialData$Genre,
                          `NA Sales`=InitialData$`NA Sales`,
                          `Critic Score`=InitialData$`Critic Score`,
                          `Critic Count`=InitialData$`Critic Count`)
#Europe
UserBasedDataEU <- tibble(Title=InitialData$Title,
                          Publisher=InitialData$Publisher,
                          Developer=InitialData$Developer,
                          `Year of Release`=InitialData$`Year of Release`,
                          Platform=InitialData$Platform,
                          Rating=InitialData$Rating,
                          Genre=InitialData$Genre,
                          `EU Sales`=InitialData$`EU Sales`,
                          `User Score`=InitialData$`User Score`,
                          `User Count`=InitialData$`User Count`)
CriticBasedDataEU <- tibble(Title=InitialData$Title,
                          Publisher=InitialData$Publisher,
                          Developer=InitialData$Developer,
                          `Year of Release`=InitialData$`Year of Release`,
                          Platform=InitialData$Platform,
                          Rating=InitialData$Rating,
                          Genre=InitialData$Genre,
                          `EU Sales`=InitialData$`EU Sales`,
                          `Critic Score`=InitialData$`Critic Score`,
                          `Critic Count`=InitialData$`Critic Count`)
#Japan
UserBasedDataJP <- tibble(Title=InitialData$Title,
                          Publisher=InitialData$Publisher,
                          Developer=InitialData$Developer,
                          `Year of Release`=InitialData$`Year of Release`,
```

```
                        Platform=InitialData$Platform,
                        Rating=InitialData$Rating,
                        Genre=InitialData$Genre,
                        `JP Sales`=InitialData$`JP Sales`,
                        `User Score`=InitialData$`User Score`,
                        `User Count`=InitialData$`User Count`)
CriticBasedDataJP <- tibble(Title=InitialData$Title,
                        Publisher=InitialData$Publisher,
                        Developer=InitialData$Developer,
                        `Year of Release`=InitialData$`Year of Release`,
                        Platform=InitialData$Platform,
                        Rating=InitialData$Rating,
                        Genre=InitialData$Genre,
                        `JP Sales`=InitialData$`JP Sales`,
                        `Critic Score`=InitialData$`Critic Score`,
                        `Critic Count`=InitialData$`Critic Count`)
#Global
UserBasedDataGlobal <- tibble(Title=InitialData$Title,
                        Publisher=InitialData$Publisher,
                        Developer=InitialData$Developer,
                        `Year of Release`=InitialData$`Year of Release`,
                        Platform=InitialData$Platform,
                        Rating=InitialData$Rating,
                        Genre=InitialData$Genre,
                        `Global Sales`=InitialData$`Global Sales`,
                        `User Score`=InitialData$`User Score`,
                        `User Count`=InitialData$`User Count`)
CriticBasedDataGlobal<- tibble(Title=InitialData$Title,
                        Publisher=InitialData$Publisher,
                        Developer=InitialData$Developer,
                        `Year of Release`=InitialData$`Year of Release`,
                        Platform=InitialData$Platform,
                        Rating=InitialData$Rating,
                        Genre=InitialData$Genre,
                        `Global Sales`=InitialData$`Global Sales`,
                        `Critic Score`=InitialData$`Critic Score`,
                        `Critic Count`=InitialData$`Critic Count`)
```

Summarization

Using the summary function, We can outline some general descriptive statistics found for each of the tables we've created.

North America

```
summary(UserBasedDataNA[,5:10])
```

```
##     Platform        Rating              Genre         NA Sales
##  PS2    :1161   T      :2420   Action      :1677   Min.   : 0.0000
##  X360   : 881   E      :2118   Sports      : 973   1st Qu.: 0.0600
##  PS3    : 790   M      :1459   Shooter     : 886   Median : 0.1500
##  PC     : 703   E10+   : 946   Role-Playing: 721   Mean   : 0.3893
##  XB     : 581          :  70   Racing      : 598   3rd Qu.: 0.3900
##  Wii    : 492   RP     :   2   Platform    : 407   Max.   :41.3600
##  (Other):2409   (Other):   2   (Other)     :1755
##    User Score     User Count
##  7.8    : 298   Min.   :    4.0
##  8      : 267   1st Qu.:   11.0
##  8.2    : 267   Median :   27.0
##  8.5    : 245   Mean   :  173.4
##  7.5    : 240   3rd Qu.:   89.0
##  7.9    : 240   Max.   :10665.0
##  (Other):5460
```

```
summary(CriticBasedDataNA[,5:10])
```

```
##     Platform        Rating              Genre         NA Sales
##  PS2    :1161   T      :2420   Action      :1677   Min.   : 0.0000
##  X360   : 881   E      :2118   Sports      : 973   1st Qu.: 0.0600
##  PS3    : 790   M      :1459   Shooter     : 886   Median : 0.1500
##  PC     : 703   E10+   : 946   Role-Playing: 721   Mean   : 0.3893
##  XB     : 581          :  70   Racing      : 598   3rd Qu.: 0.3900
##  Wii    : 492   RP     :   2   Platform    : 407   Max.   :41.3600
##  (Other):2409   (Other):   2   (Other)     :1755
##   Critic Score   Critic Count
##  8      : 231   Min.   :  3.00
##  7.5    : 223   1st Qu.: 14.00
##  7.8    : 221   Median : 24.00
##  7.3    : 219   Mean   : 28.78
##  7.1    : 218   3rd Qu.: 39.00
##  7.7    : 215   Max.   :113.00
##  (Other):5690
```

Europe

```
summary(UserBasedDataEU[,5:10])
```

```
##     Platform        Rating              Genre         EU Sales
##  PS2    :1161   T      :2420   Action      :1677   Min.   : 0.0000
##  X360   : 881   E      :2118   Sports      : 973   1st Qu.: 0.0200
##  PS3    : 790   M      :1459   Shooter     : 886   Median : 0.0600
##  PC     : 703   E10+   : 946   Role-Playing: 721   Mean   : 0.2331
##  XB     : 581          :  70   Racing      : 598   3rd Qu.: 0.2100
##  Wii    : 492   RP     :   2   Platform    : 407   Max.   :28.9600
##  (Other):2409   (Other):   2   (Other)     :1755
##    User Score     User Count
##  7.8    : 298   Min.   :    4.0
##  8      : 267   1st Qu.:   11.0
##  8.2    : 267   Median :   27.0
##  8.5    : 245   Mean   :  173.4
##  7.5    : 240   3rd Qu.:   89.0
##  7.9    : 240   Max.   :10665.0
##  (Other):5460
```

```
summary(CriticBasedDataEU[,5:10])
```

```
##     Platform        Rating              Genre         EU Sales
##  PS2    :1161   T      :2420   Action      :1677   Min.   : 0.0000
##  X360   : 881   E      :2118   Sports      : 973   1st Qu.: 0.0200
##  PS3    : 790   M      :1459   Shooter     : 886   Median : 0.0600
##  PC     : 703   E10+   : 946   Role-Playing: 721   Mean   : 0.2331
##  XB     : 581          :  70   Racing      : 598   3rd Qu.: 0.2100
##  Wii    : 492   RP     :   2   Platform    : 407   Max.   :28.9600
##  (Other):2409   (Other):   2   (Other)     :1755
##   Critic Score   Critic Count
##  8      : 231   Min.   :  3.00
##  7.5    : 223   1st Qu.: 14.00
##  7.8    : 221   Median : 24.00
##  7.3    : 219   Mean   : 28.78
##  7.1    : 218   3rd Qu.: 39.00
##  7.7    : 215   Max.   :113.00
##  (Other):5690
```

Japan

```
summary(UserBasedDataJP[,5:10])
```

```
##     Platform        Rating              Genre         JP Sales
##  PS2    :1161   T      :2420   Action      :1677   Min.   :0.00000
##  X360   : 881   E      :2118   Sports      : 973   1st Qu.:0.00000
##  PS3    : 790   M      :1459   Shooter     : 886   Median :0.00000
##  PC     : 703   E10+   : 946   Role-Playing: 721   Mean   :0.06295
##  XB     : 581          :  70   Racing      : 598   3rd Qu.:0.01000
##  Wii    : 492   RP     :   2   Platform    : 407   Max.   :6.50000
##  (Other):2409   (Other):   2   (Other)     :1755
##    User Score     User Count
##  7.8    : 298   Min.   :    4.0
##  8      : 267   1st Qu.:   11.0
##  8.2    : 267   Median :   27.0
##  8.5    : 245   Mean   :  173.4
##  7.5    : 240   3rd Qu.:   89.0
##  7.9    : 240   Max.   :10665.0
##  (Other):5460
```

```
summary(CriticBasedDataJP[,5:10])
```

```
##     Platform        Rating              Genre         JP Sales
##  PS2    :1161   T      :2420   Action      :1677   Min.   :0.00000
##  X360   : 881   E      :2118   Sports      : 973   1st Qu.:0.00000
##  PS3    : 790   M      :1459   Shooter     : 886   Median :0.00000
##  PC     : 703   E10+   : 946   Role-Playing: 721   Mean   :0.06295
##  XB     : 581          :  70   Racing      : 598   3rd Qu.:0.01000
##  Wii    : 492   RP     :   2   Platform    : 407   Max.   :6.50000
##  (Other):2409   (Other):   2   (Other)     :1755
##   Critic Score   Critic Count
##  8      : 231   Min.   :  3.00
##  7.5    : 223   1st Qu.: 14.00
##  7.8    : 221   Median : 24.00
##  7.3    : 219   Mean   : 28.78
##  7.1    : 218   3rd Qu.: 39.00
##  7.7    : 215   Max.   :113.00
##  (Other):5690
```

Global

```
summary(UserBasedDataGlobal[,5:10])
```

```
##     Platform        Rating              Genre       Global Sales
##  PS2    :1161   T      :2420   Action      :1677   Min.   : 0.0100
##  X360   : 881   E      :2118   Sports      : 973   1st Qu.: 0.1100
##  PS3    : 790   M      :1459   Shooter     : 886   Median : 0.2900
##  PC     : 703   E10+   : 946   Role-Playing: 721   Mean   : 0.7671
##  XB     : 581          :  70   Racing      : 598   3rd Qu.: 0.7500
##  Wii    : 492   RP     :   2   Platform    : 407   Max.   :82.5300
##  (Other):2409   (Other):   2   (Other)     :1755
##    User Score     User Count
##  7.8    : 298   Min.   :    4.0
##  8      : 267   1st Qu.:   11.0
##  8.2    : 267   Median :   27.0
##  8.5    : 245   Mean   :  173.4
##  7.5    : 240   3rd Qu.:   89.0
##  7.9    : 240   Max.   :10665.0
##  (Other):5460
```

```
summary(CriticBasedDataGlobal[,5:10])
```

```
##     Platform        Rating              Genre       Global Sales
##  PS2    :1161   T      :2420   Action      :1677   Min.   : 0.0100
##  X360   : 881   E      :2118   Sports      : 973   1st Qu.: 0.1100
##  PS3    : 790   M      :1459   Shooter     : 886   Median : 0.2900
##  PC     : 703   E10+   : 946   Role-Playing: 721   Mean   : 0.7671
##  XB     : 581          :  70   Racing      : 598   3rd Qu.: 0.7500
##  Wii    : 492   RP     :   2   Platform    : 407   Max.   :82.5300
##  (Other):2409   (Other):   2   (Other)     :1755
##   Critic Score   Critic Count
##  8      : 231   Min.   :  3.00
##  7.5    : 223   1st Qu.: 14.00
##  7.8    : 221   Median : 24.00
##  7.3    : 219   Mean   : 28.78
##  7.1    : 218   3rd Qu.: 39.00
##  7.7    : 215   Max.   :113.00
##  (Other):5690
```

However, to get into more specifics, we created tables that better show the relationship between sales figures and scores. In addition, these tables allows easier analysis towards a game's relationship between its developer, publisher, and rating with regards to its sales.

North America

```
SalesVsUserScoreNA <- UserBasedDataNA[order(-UserBasedDataNA$`NA Sales`),] #Organizes list by sales
SalesVsUserScoreNA <- head(SalesVsUserScoreNA, n=10)
SalesVsCriticScoreNA <- CriticBasedDataNA[order(-CriticBasedDataNA$`NA Sales`),] #Organizes list by sales
SalesVsCriticScoreNA <- head(SalesVsCriticScoreNA, n=10)
show(SalesVsUserScoreNA)
```

```
## # A tibble: 10 x 10
##    Title Publisher Developer `Year of Releas~ Platform Rating Genre
##    <fct> <fct>     <fct>     <fct>            <fct>    <fct>  <fct>
##  1 Wii ~ Nintendo  Nintendo  2006             Wii      E      Spor~
##  2 Mari~ Nintendo  Nintendo  2008             Wii      E      Raci~
##  3 Wii ~ Nintendo  Nintendo  2009             Wii      E      Spor~
##  4 Kine~ Microsof~ Good Sci~ 2010             X360     E      Misc
##  5 New ~ Nintendo  Nintendo  2009             Wii      E      Plat~
##  6 Wii ~ Nintendo  Nintendo  2006             Wii      E      Misc
##  7 New ~ Nintendo  Nintendo  2006             DS       E      Plat~
##  8 Mari~ Nintendo  Nintendo  2005             DS       E      Raci~
##  9 Call~ Activisi~ Treyarch  2010             X360     M      Shoo~
## 10 Gran~ Take-Two~ Rockstar~ 2013             X360     M      Acti~
## # ... with 3 more variables: `NA Sales` <dbl>, `User Score` <fct>, `User
## #   Count` <int>
```

```
show(SalesVsCriticScoreNA)
```

```
## # A tibble: 10 x 10
##    Title Publisher Developer `Year of Releas~ Platform Rating Genre
##    <fct> <fct>     <fct>     <fct>            <fct>    <fct>  <fct>
##  1 Wii ~ Nintendo  Nintendo  2006             Wii      E      Spor~
##  2 Mari~ Nintendo  Nintendo  2008             Wii      E      Raci~
##  3 Wii ~ Nintendo  Nintendo  2009             Wii      E      Spor~
##  4 Kine~ Microsof~ Good Sci~ 2010             X360     E      Misc
##  5 New ~ Nintendo  Nintendo  2009             Wii      E      Plat~
##  6 Wii ~ Nintendo  Nintendo  2006             Wii      E      Misc
##  7 New ~ Nintendo  Nintendo  2006             DS       E      Plat~
##  8 Mari~ Nintendo  Nintendo  2005             DS       E      Raci~
##  9 Call~ Activisi~ Treyarch  2010             X360     M      Shoo~
## 10 Gran~ Take-Two~ Rockstar~ 2013             X360     M      Acti~
## # ... with 3 more variables: `NA Sales` <dbl>, `Critic Score` <fct>,
## #   `Critic Count` <int>
```

Europe

```
SalesVsUserScoreEU <- UserBasedDataEU[order(-UserBasedDataEU$`EU Sales`),] #Organizes table by sales
SalesVsUserScoreEU <- head(SalesVsUserScoreEU, n=10)
SalesVsCriticScoreEU <- CriticBasedDataEU[order(-CriticBasedDataEU$`EU Sales`),] #Organizes list by sales
SalesVsCriticScoreEU <- head(SalesVsCriticScoreEU, n=10)
show(SalesVsUserScoreEU)
```

```
## # A tibble: 10 x 10
##    Title Publisher Developer `Year of Releas~ Platform Rating Genre
##    <fct> <fct>     <fct>     <fct>            <fct>    <fct>  <fct>
##  1 Wii ~ Nintendo  Nintendo  2006             Wii      E      Spor~
##  2 Mari~ Nintendo  Nintendo  2008             Wii      E      Raci~
##  3 Wii ~ Nintendo  Nintendo  2009             Wii      E      Spor~
##  4 Brai~ Nintendo  Nintendo  2005             DS       E      Misc
##  5 Wii ~ Nintendo  Nintendo  2006             Wii      E      Misc
##  6 New ~ Nintendo  Nintendo  2006             DS       E      Plat~
##  7 Gran~ Take-Two~ Rockstar~ 2013             PS3      M      Acti~
##  8 Wii ~ Nintendo  Nintendo  2009             Wii      E      Spor~
##  9 Wii ~ Nintendo  Nintendo  2007             Wii      E      Spor~
## 10 Mari~ Nintendo  Nintendo  2005             DS       E      Raci~
## # ... with 3 more variables: `EU Sales` <dbl>, `User Score` <fct>, `User
## #   Count` <int>
```

```
show(SalesVsCriticScoreEU)
```

```
## # A tibble: 10 x 10
##    Title Publisher Developer `Year of Releas~ Platform Rating Genre
##    <fct> <fct>     <fct>     <fct>            <fct>    <fct>  <fct>
##  1 Wii ~ Nintendo  Nintendo  2006             Wii      E      Spor~
##  2 Mari~ Nintendo  Nintendo  2008             Wii      E      Raci~
##  3 Wii ~ Nintendo  Nintendo  2009             Wii      E      Spor~
##  4 Brai~ Nintendo  Nintendo  2005             DS       E      Misc
##  5 Wii ~ Nintendo  Nintendo  2006             Wii      E      Misc
##  6 New ~ Nintendo  Nintendo  2006             DS       E      Plat~
##  7 Gran~ Take-Two~ Rockstar~ 2013             PS3      M      Acti~
##  8 Wii ~ Nintendo  Nintendo  2009             Wii      E      Spor~
##  9 Wii ~ Nintendo  Nintendo  2007             Wii      E      Spor~
## 10 Mari~ Nintendo  Nintendo  2005             DS       E      Raci~
## # ... with 3 more variables: `EU Sales` <dbl>, `Critic Score` <fct>,
## #   `Critic Count` <int>
```

Japan

```
SalesVsUserScoreJP <- UserBasedDataJP[order(-UserBasedDataJP$`JP Sales`),] #Organizes list by sales
SalesVsUserScoreJP <- head(SalesVsUserScoreJP, n=10)
SalesVsCriticScoreJP <- CriticBasedDataJP[order(-CriticBasedDataJP$`JP Sales`),] #Organizes list by sales
SalesVsCriticScoreJP <- head(SalesVsCriticScoreJP, n=10)
show(SalesVsUserScoreJP)
```

```
## # A tibble: 10 x 10
##    Title Publisher Developer `Year of Releas~ Platform Rating Genre
##    <fct> <fct>     <fct>     <fct>            <fct>    <fct>  <fct>
##  1 New ~ Nintendo  Nintendo  2006             DS       E      Plat~
##  2 Anim~ Nintendo  Nintendo  2005             DS       E      Simu~
##  3 Brai~ Nintendo  Nintendo  2005             DS       E      Puzz~
##  4 New ~ Nintendo  Nintendo  2009             Wii      E      Plat~
##  5 Anim~ Nintendo  Nintendo  2012             3DS      E      Simu~
##  6 Drag~ Nintendo  Level 5   2009             DS       E10+   Role~
##  7 Brai~ Nintendo  Nintendo  2005             DS       E      Misc
##  8 Mari~ Nintendo  Nintendo  2005             DS       E      Raci~
##  9 Mons~ Capcom    Capcom    2008             PSP      T      Role~
## 10 Mari~ Nintendo  Nintendo  2008             Wii      E      Raci~
## # ... with 3 more variables: `JP Sales` <dbl>, `User Score` <fct>, `User
## #   Count` <int>
```

```
show(SalesVsCriticScoreJP)
```

```
## # A tibble: 10 x 10
##    Title Publisher Developer `Year of Releas~ Platform Rating Genre
##    <fct> <fct>     <fct>     <fct>            <fct>    <fct>  <fct>
##  1 New ~ Nintendo  Nintendo  2006             DS       E      Plat~
##  2 Anim~ Nintendo  Nintendo  2005             DS       E      Simu~
##  3 Brai~ Nintendo  Nintendo  2005             DS       E      Puzz~
##  4 New ~ Nintendo  Nintendo  2009             Wii      E      Plat~
##  5 Anim~ Nintendo  Nintendo  2012             3DS      E      Simu~
##  6 Drag~ Nintendo  Level 5   2009             DS       E10+   Role~
##  7 Brai~ Nintendo  Nintendo  2005             DS       E      Misc
##  8 Mari~ Nintendo  Nintendo  2005             DS       E      Raci~
##  9 Mons~ Capcom    Capcom    2008             PSP      T      Role~
## 10 Mari~ Nintendo  Nintendo  2008             Wii      E      Raci~
## # ... with 3 more variables: `JP Sales` <dbl>, `Critic Score` <fct>,
## #   `Critic Count` <int>
```

Global

```
SalesVsUserScoreGlobal <- UserBasedDataGlobal[order(-UserBasedDataGlobal$`Global Sales`),] #Organizes list b
y sales
SalesVsUserScoreGlobal <- head(SalesVsUserScoreGlobal, n=10)
SalesVsCriticScoreGlobal <- CriticBasedDataGlobal[order(-CriticBasedDataGlobal$`Global Sales`),] #Organizes
list by sales
SalesVsCriticScoreGlobal <- head(SalesVsCriticScoreGlobal, n=10)
show(SalesVsUserScoreGlobal)
```

```
## # A tibble: 10 x 10
##    Title Publisher Developer `Year of Releas~ Platform Rating Genre
##    <fct> <fct>     <fct>     <fct>            <fct>    <fct>  <fct>
##  1 Wii ~ Nintendo  Nintendo  2006             Wii      E      Spor~
##  2 Mari~ Nintendo  Nintendo  2008             Wii      E      Raci~
##  3 Wii ~ Nintendo  Nintendo  2009             Wii      E      Spor~
##  4 New ~ Nintendo  Nintendo  2006             DS       E      Plat~
##  5 Wii ~ Nintendo  Nintendo  2006             Wii      E      Misc
##  6 New ~ Nintendo  Nintendo  2009             Wii      E      Plat~
##  7 Mari~ Nintendo  Nintendo  2005             DS       E      Raci~
##  8 Wii ~ Nintendo  Nintendo  2007             Wii      E      Spor~
##  9 Kine~ Microsof~ Good Sci~ 2010             X360     E      Misc
## 10 Wii ~ Nintendo  Nintendo  2009             Wii      E      Spor~
## # ... with 3 more variables: `Global Sales` <dbl>, `User Score` <fct>,
## #   `User Count` <int>
```

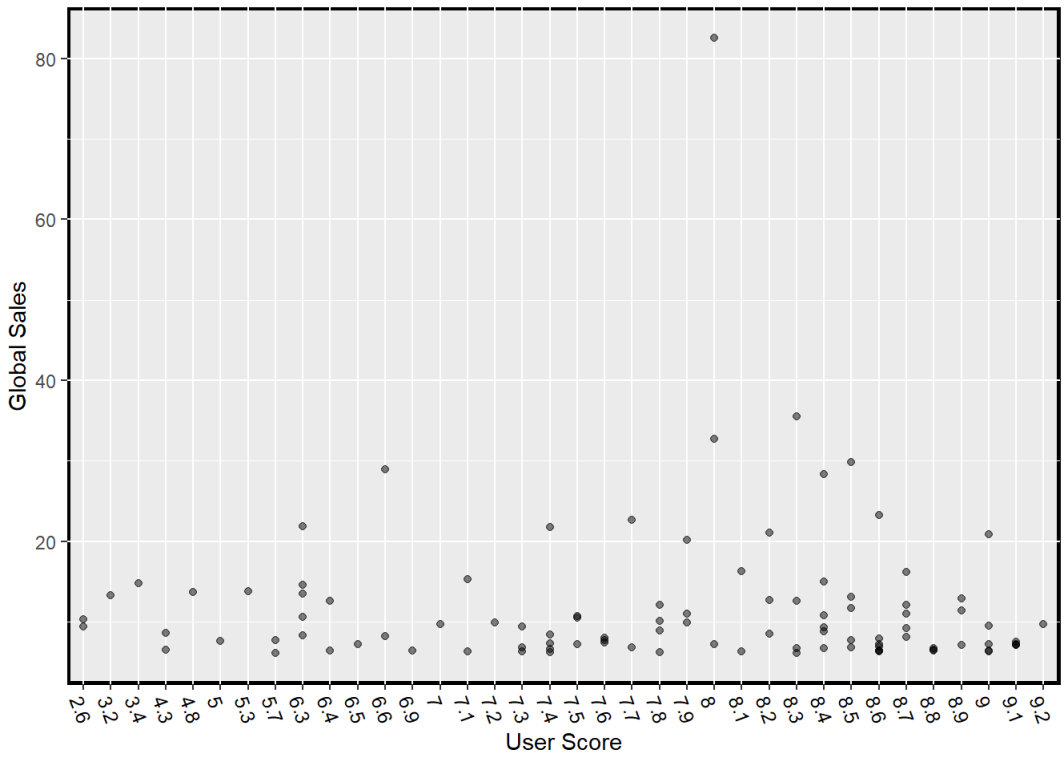
```
show(SalesVsCriticScoreGlobal)
```

```
## # A tibble: 10 x 10
##    Title Publisher Developer `Year of Releas~ Platform Rating Genre
##    <fct> <fct>     <fct>     <fct>            <fct>    <fct>  <fct>
##  1 Wii ~ Nintendo  Nintendo  2006             Wii      E      Spor~
##  2 Mari~ Nintendo  Nintendo  2008             Wii      E      Raci~
##  3 Wii ~ Nintendo  Nintendo  2009             Wii      E      Spor~
##  4 New ~ Nintendo  Nintendo  2006             DS       E      Plat~
##  5 Wii ~ Nintendo  Nintendo  2006             Wii      E      Misc
##  6 New ~ Nintendo  Nintendo  2009             Wii      E      Plat~
##  7 Mari~ Nintendo  Nintendo  2005             DS       E      Raci~
##  8 Wii ~ Nintendo  Nintendo  2007             Wii      E      Spor~
##  9 Kine~ Microsof~ Good Sci~ 2010             X360     E      Misc
## 10 Wii ~ Nintendo  Nintendo  2009             Wii      E      Spor~
## # ... with 3 more variables: `Global Sales` <dbl>, `Critic Score` <fct>,
## #   `Critic Count` <int>
```

To aid in visualization, we can use graphs to help show the relationship between scores, ratings, and publishers with regards to sales.

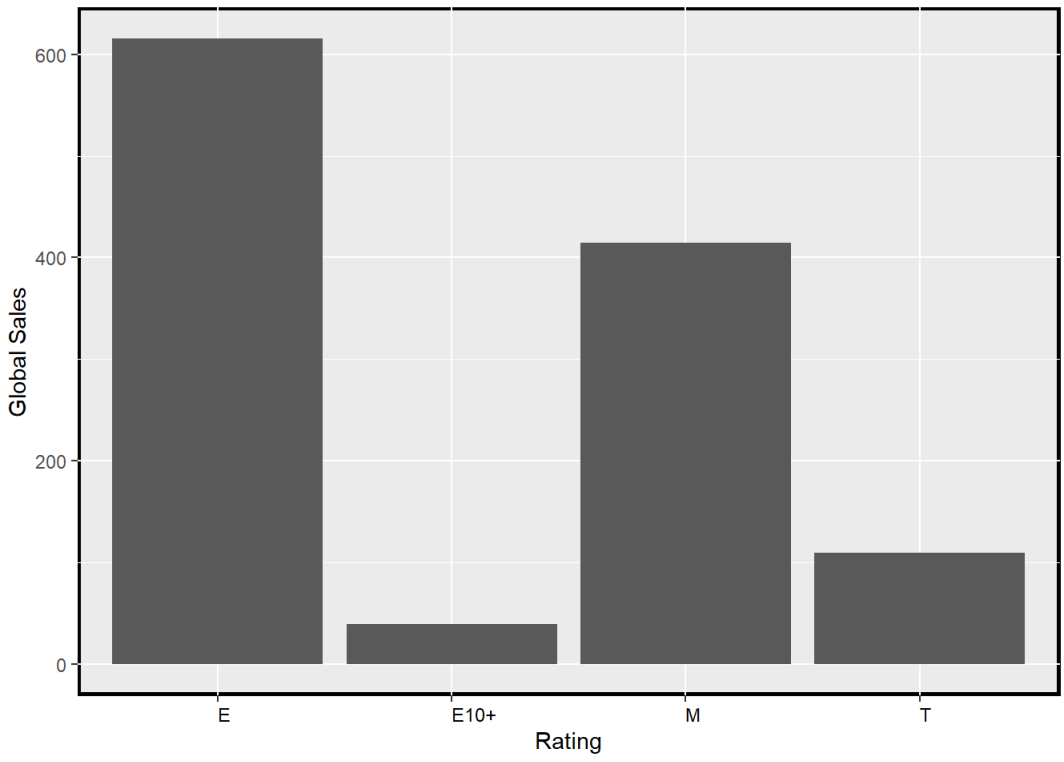
Scores vs Sales

```
GraphData <- UserBasedDataGlobal[order(-UserBasedDataGlobal$`Global Sales`),]
GraphData <- head(GraphData, n=100)
ggplot(data = GraphData, mapping=aes(x=`User Score`, y=`Global Sales`)) +
  geom_point(alpha=1/2) +
  theme(panel.background = element_rect(color = "black" , size = 1.5), axis.text.x = element_text(size=9, an
gle = 290, vjust = 1, hjust = 0, color = "black"))
```



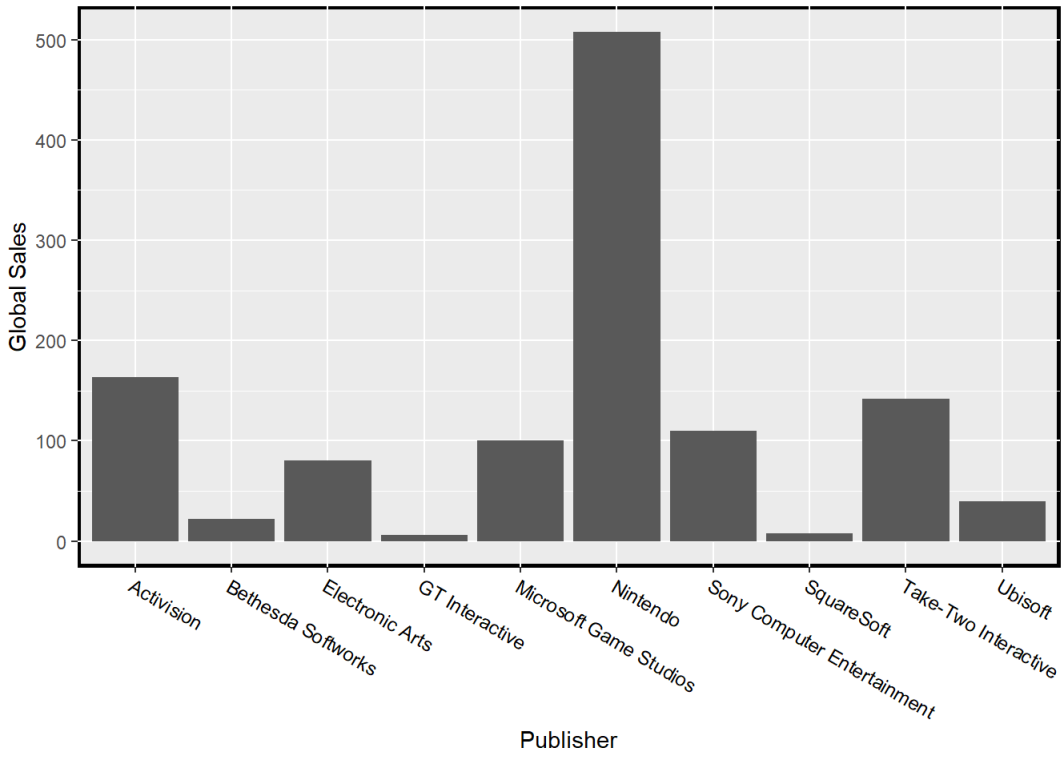
Ratings vs Sales

```
ggplot(data = GraphData, mapping=aes(x=`Rating`, y=`Global Sales`)) +
  geom_col() +
  theme(panel.background = element_rect(color = "black" , size = 1.5), axis.text.x = element_text(size=9, vj
ust = 1, hjust = 0, color = "black"))
```



Publishers vs Sales

```
ggplot(data = GraphData, mapping=aes(x=`Publisher`, y=`Global Sales`)) +
  geom_col() +
  theme(panel.background = element_rect(color = "black" , size = 1.5), axis.text.x = element_text(size=9, an
gle = 330, vjust = 1, hjust = 0, color = "black"))
```



Research Questions

1. Does a higher age rating for a game have any kind of negative/positive effects on a game's sales? If so, can we determine some kind of margin that higher rated games will have in terms of success/failure compared to E rated games? I find this one interesting because, in theory, being a lower rated game would mean you would have access to a larger audience. This seems to be true for the data we have within our top ten tables, but we can see a small indication that the higher rated games become more frequent after the first seven or so top selling games, especially in North America.
2. Based on each console generation, what games sold the best/scored the best and was there any indication that some publishers/developers did better or worse in some generations? Have there been any changes in reviewer trends? I find this one compelling because as we go into a new generation for games, audience expectations and taste will begin to differ. Pursuing a question like this will help show any trends reviews have gone through as each new generation comes to pass. We can also track the success/failure a publisher/developer has received as the years go on