Earning in Esports

Introduction

Video game competitions are known as esports. Esports usually take the form of organized competitions for multiplayer video games, with a focus on professional players playing individually or in teams. Over the past ten years, esports have experienced tremendous growth and increased popularity. This report includes specifics on the earnings generated by esports. The information came from Kaggle.The dataset includes information on earnings from various game types from 1989 to 2020. The dataset includes earnings information for several games in addition to information on genre, top countrnhy, release year, and other factors. For this project, our primary goal is to concentrate on the total revenue generated by diverse game genres.

Data Cleaning

```
#Summarzing the esport data to find summary of diiferent fields to find odd and unwanted data
summary(esports)
```

```
##       IdNo         TotalMoney          GameName            Genre
##  Min.   :151.0   Min.   :        0   Length:504         Length:504
##  1st Qu.:276.8   1st Qu.:     1375   Class :character   Class :character
##  Median :454.5   Median :    36009   Mode  :character   Mode  :character
##  Mean   :441.0   Mean   :  1744743
##  3rd Qu.:580.2   3rd Qu.:   205752
##  Max.   :706.0   Max.   :227419905
##     PlayerNo        TournamentNo     Top_Country        Top_Country_Earnings
##  Min.   :    0.0   Min.   :   0.00   Length:504         Min.   :       0
##  1st Qu.:    4.0   1st Qu.:   1.00   Class :character   1st Qu.:     391
##  Median :   20.0   Median :   5.00   Mode  :character   Median :   17117
##  Mean   :  186.7   Mean   :  80.75                      Mean   :  601082
##  3rd Qu.:   86.0   3rd Qu.:  25.25                      3rd Qu.:  103604
##  Max.   :13032.0   Max.   :5909.00                      Max.   :65060611
##   Releaseyear
##  Min.   :  11
##  1st Qu.:2006
##  Median :2013
##  Mean   :2007
##  3rd Qu.:2017
##  Max.   :2020
```

```
#Removing duplicated data if their is any
esports <- esports[!duplicated(esports), ]

#Since we are looking for total earning for different game we will remove all the data that has total e
#since it doesnt match up with out objectives
#Removing the data that has total money earned 0
esports <- esports[esports$TotalMoney != 0, ]

#addition we can find there are mutiple data with top country with none value as it is not useful for o
```

```r
esports <- esports[esports$Top_Country != "None", ]

#The variable PlayerNo and Top_Country_Earning also has data with 0 so we removing those as well
esports <- esports[esports$Top_Country_Earnings != 0, ]

#Releaseyear we can find that one of the data has 11 which doesnt make any sense
#since it is a outlier and the data is of time period of 1989 to 2020 we can presume the data to be 201
esports$Releaseyear[esports$Releaseyear == 11] <- 2011

head(esports)
```

```
##   IdNo TotalMoney                 GameName         Genre PlayerNo TournamentNo
## 1  613     310.00 Acceleration of SUGURI 2 Fighting Game        6            2
## 2  510  190075.61           Age of Empires      Strategy      246           95
## 3  179 1492489.51        Age of Empires II      Strategy      956          576
## 4  177   68113.85       Age of Empires III      Strategy      106           76
## 5  540    2668.00    Age of Empires Online      Strategy       16            7
## 6  178   52360.00         Age of Mythology      Strategy       22            6
##                 Top_Country Top_Country_Earnings Releaseyear
## 1             United States               310.00        2018
## 2                 Viet Nam            123408.43        1997
## 3                     China            174252.44        1999
## 4             United States             18523.52        2005
## 5                   Germany              1264.00        2011
## 6 Taiwan, Republic of China             22000.00        2002
```

```r
summary(esports)
```

```
##       IdNo         TotalMoney          GameName            Genre
##  Min.   :151.0   Min.   :       10   Length:444         Length:444
##  1st Qu.:274.8   1st Qu.:     4457   Class :character   Class :character
##  Median :455.5   Median :    50000   Mode  :character   Mode  :character
##  Mean   :442.4   Mean   :  1978937
##  3rd Qu.:587.2   3rd Qu.:   283092
##  Max.   :706.0   Max.   :227419905
##     PlayerNo        TournamentNo     Top_Country        Top_Country_Earnings
##  Min.   :    1.0   Min.   :   1.00   Length:444         Min.   :      10
##  1st Qu.:    7.0   1st Qu.:   2.00   Class :character   1st Qu.:    1819
##  Median :   31.5   Median :   7.00   Mode  :character   Median :   24528
##  Mean   :  211.8   Mean   :  91.49                      Mean   :  682310
##  3rd Qu.:  101.8   3rd Qu.:  35.00                      3rd Qu.:  153137
##  Max.   :13032.0   Max.   :5909.00                      Max.   :65060611
##   Releaseyear
##  Min.   :1989
##  1st Qu.:2006
##  Median :2013
##  Mean   :2011
##  3rd Qu.:2017
##  Max.   :2020
```

In order to clean the data, I started by looking over the table's summary and finding some data points that didn't seem right for our analysis. We will eliminate all the data that has a total earning of zero because we are trying to determine the top earning game genre and are seeking for total earning for distinct games.Furthermore, we discovered that several data points for the top country have no value. Since these data points are useless for our purposes, I eliminated all of the zero-valued data points. I also eliminated the

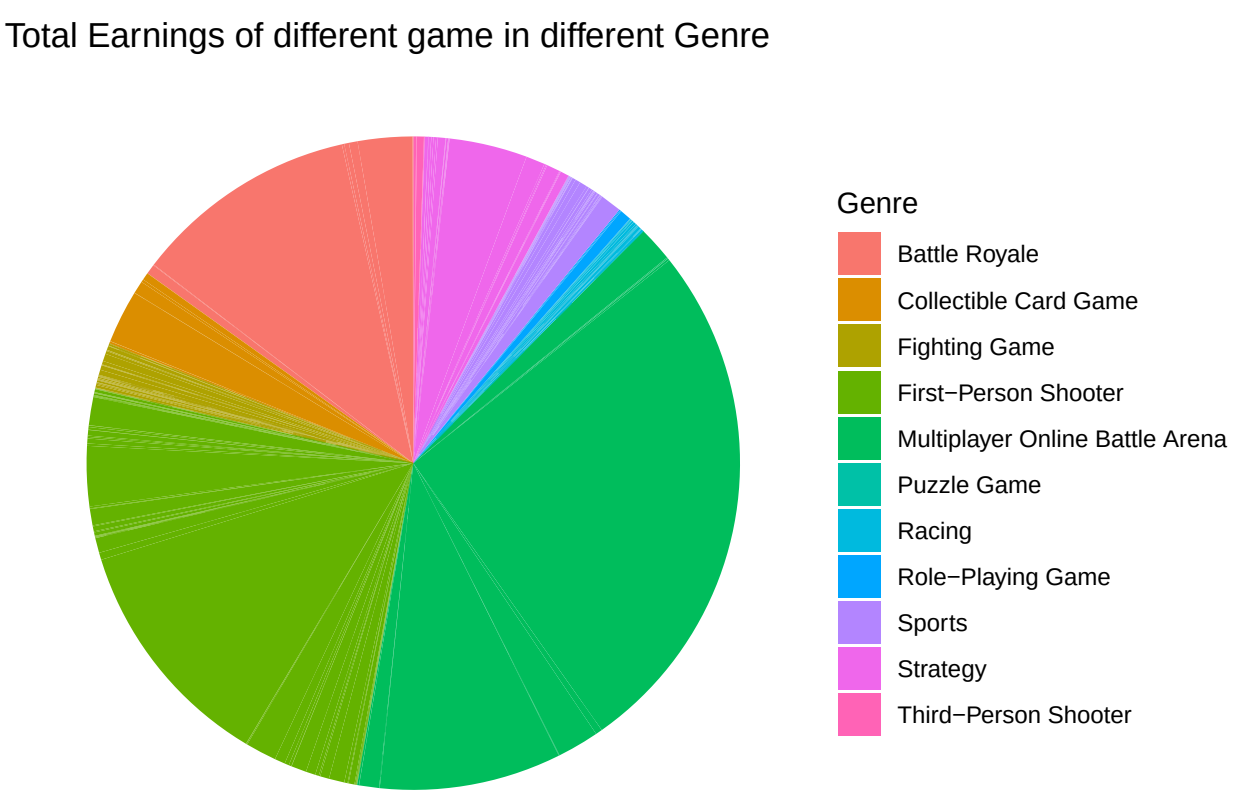
variables PlayerNo and Top_Country_Earning since they too had data that was zero. I updated the data to 2011 because it is the only year with 11 in the period from 1989 to 2020 because one of the data sets had 11 for a year that felt out of order.

Awful Representation

Pie charts is my choice for awful data presentations. This is a very poor representation of the data since it lacks relevant information that the user can use. The color palette chosen to depict several genres as similar color gradients makes it more difficult to distinguish between different sets of data. Even if the graph's portions are properly scaled, it lacks a numerical value, making it impossible for us to determine the true values of those genres.

```
#Awful representation of data in piechart
pie_chart <- ggplot(esports, aes(x = "", y = TotalMoney, fill = Genre)) +
  geom_bar(stat = "identity", width = 1) +
  coord_polar("y", start = 0) +  # Convert the bar plot to a pie chart
  labs(title = "Total Earnings of different game in different Genre",
       fill = "Genre") +
  theme_void()

pie_chart
```

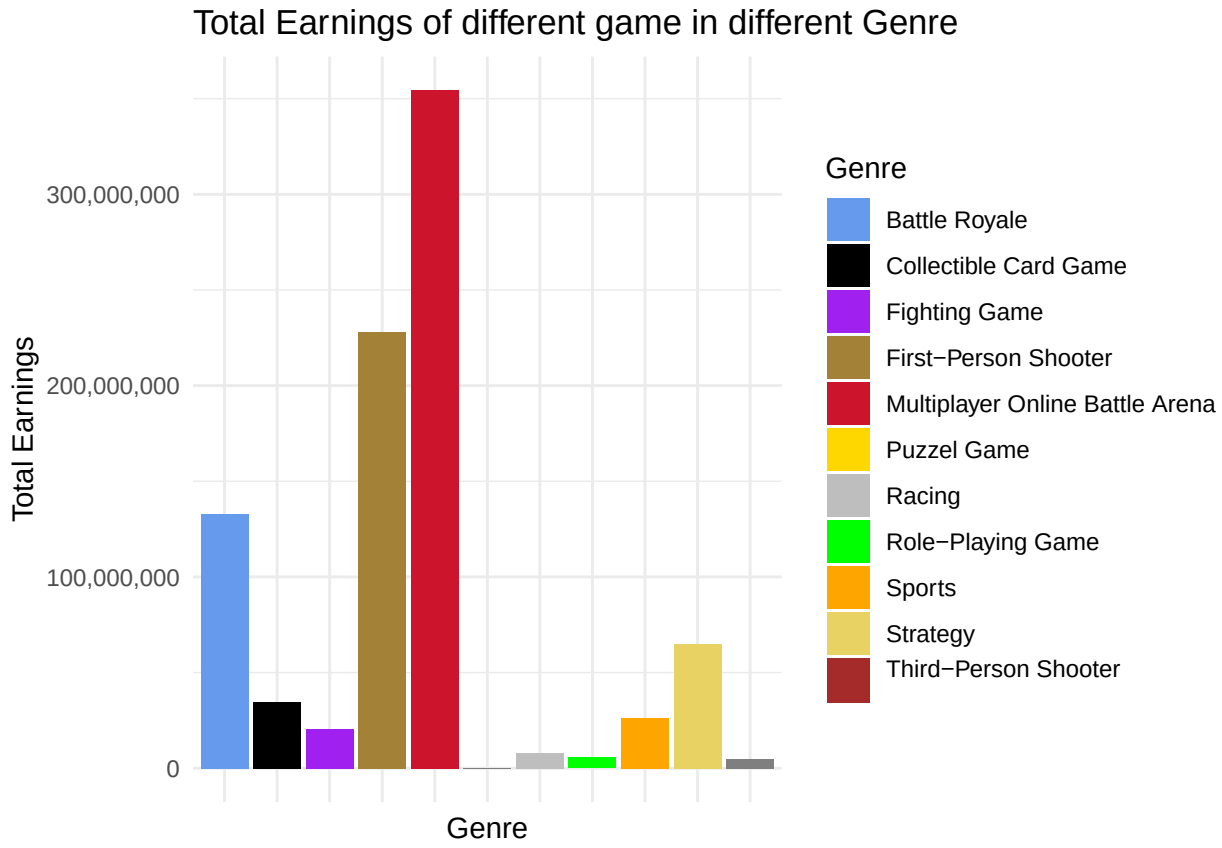


Useful representation

I believe that a bar graph is the most fundamental and easily comprehensible type of graph, thus I decided to utilize it as the helpful representation. The total earnings of all the games that are ranked in a particular genre are displayed on the barplot. With a barplot instead of a pie chart, it is easier to compare the earnings

of various genres and see the overall earnings as a numerical amount. The colors utilized in the garph are those that symbolize the well-known game genre. As a result, the hue became more striking and contrasted.

```
#Barplot for proper representation for data
ggplot(esports, aes(x = Genre, y = TotalMoney, fill = Genre)) +
  geom_bar(stat = "identity") +
  labs(title = "Total Earnings of different game in different Genre",
       x = "Genre",
       y = "Total Earnings",
       fill = "Genre") +
  theme_minimal()+
  scale_y_continuous(labels = scales::comma)+
  scale_fill_manual(values = c("Battle Royale" = "#669aed", "Collectible Card Game" = "black", "Fightin
                               "First-Person Shooter" = "#a38137", "Multiplayer Online Battle Arena" =
" = "brown"))+
  theme(axis.text.x = element_blank())
```

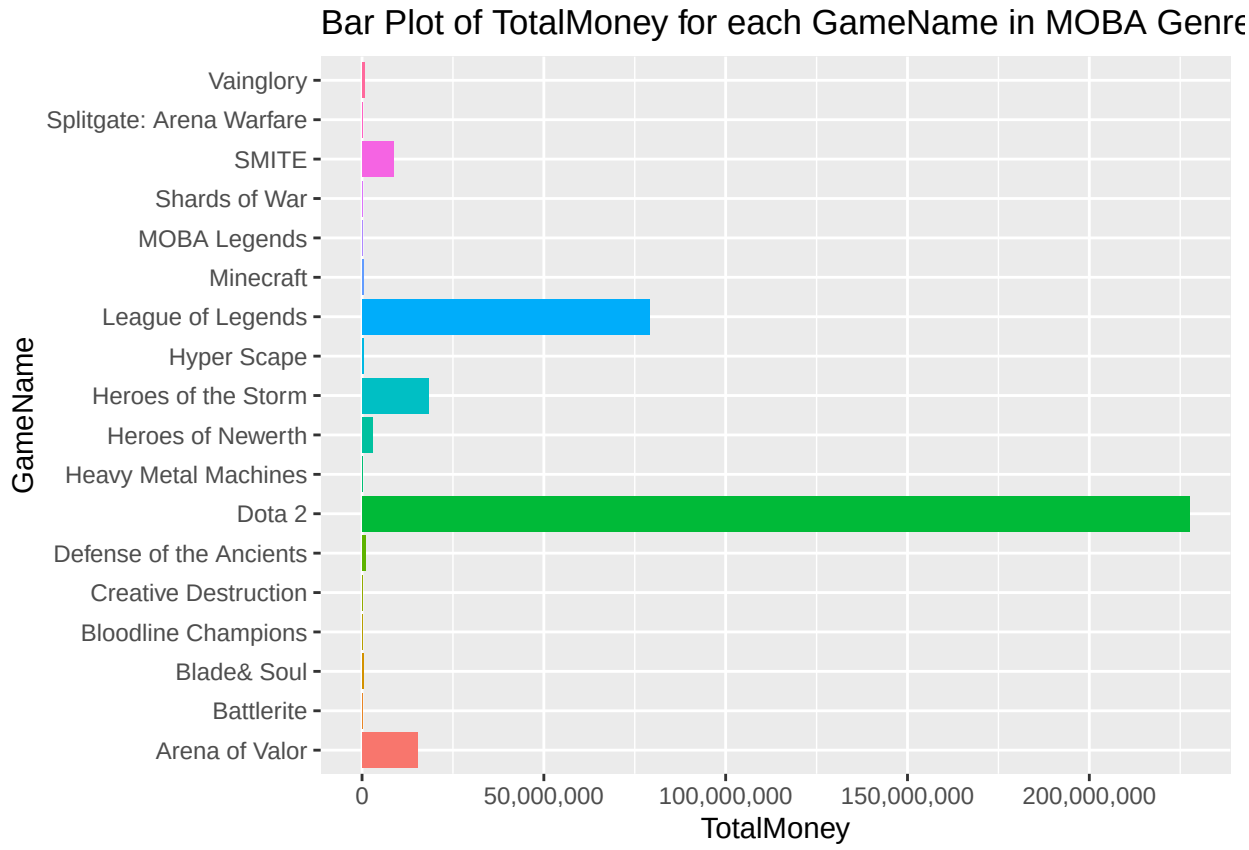


Given that our main goal was to identify the game with the highest earnings overall, I made the decision to create a second graph that would display the title with the most earnings inside the most earned genre from the first one. In order to show the game with the highest earnings, I kept using a bar plot for the final representation.

```
# Subset data for the MOBA genre
moba_data <- esports[esports$Genre == "Multiplayer Online Battle Arena", ]


# Create a scatter plot
ggplot(moba_data, aes(y = GameName, x = TotalMoney, fill = GameName)) +
```

```
  geom_bar(stat = "identity") +
  labs(title = "Bar Plot of TotalMoney for each GameName in MOBA Genre",
       y = "GameName",
       x = "TotalMoney")+
  scale_x_continuous(labels = scales::comma)+
  theme(legend.position = "none")
```



Data Visualization info

Here, the data aims to illustrate the earnings of various gaming genres. The first game uses a pie chart to display the proportion of size differences, but it does not provide any numerical or appropriate visual representation for any of the data. The second graph, a barplot, aims to depict the same data in a way that is easier to read and more quantitatively feasible.

Conclusion

With all those variable available I wanted to focus mainly on finding what Genre and which game is biggest earning in Esport games. From the data visualization we can find that Multiplayer Online Battle Arena(MOBA) and among MOBA Dota 2 is the highest earning game. So if you want to earn money through game Dota2 is the best game.

Reference

-https://www.kaggle.com/code

-https://www.kaggle.com/datasets/rushikes

-hhiray/esport-earnings

-https://en.wikipedia.org/wiki/Esports

-https://www.esportsearnings.com/games/browse-by-genre

-https://www.canva.com