

Disney movies, a money question

Jeanneret Emilie

2022-11-23



Introduction

Have you ever seen a Disney movie ? I am sure that you cannot respond “no” to that question. If you really never have seen one, I bet that you know the worldwide famous mouse (of course I am speaking of Mickey Mouse).

And this is because The Walt Disney Company is, since 2012, the first group of entertainment across the world ! Now talking a little bit about the history of this company, we know that it has been created in 1923 (100 years next year ! woaw). The founders are Roy and Walter Disney, but the latter is the main character in the Disney history.

They create many types of movie and you can easily find your way through it. Indeed, it's adapted to the whole family. The number of movies they produced is about 529, that is huge ! The list of all movie would take quite a lot of time to go through but the main element we are interested in in this research are the income of each movie, and the genre of each movie.

Research question

This research is interested in the evolution of incomings of Disney movies through time. We also want to see if there is an impact of the genre of movie on the income they engender. Of course, we will need to adapt the table with inflation rate in order to adapt our data. The main question here is **Is there an impact of the genre of Disney movies incomings over time ?**

Hypothesis

N0: There is no impact of the genre on the incoming

N1: There is an impact of the genre on the incoming

Data

The data comes from kaggle, where was hidden a very interesting dataset about disney movies. For this research, we do not need every details and every lines, even if that would be very interesting to analyze.

As we can see below, the main data table is displaying in the following way. There are the title of each movies in the first column, then the related month and year of release is displayed. There are also the revenue of the movie in dollars and the revenue but inflation adjusted. As we are not focusing on the inflation, we are not going to analyze the difference between those two columns very deeply but it might be interesting tho to have a look at it.

Table 1: Data table

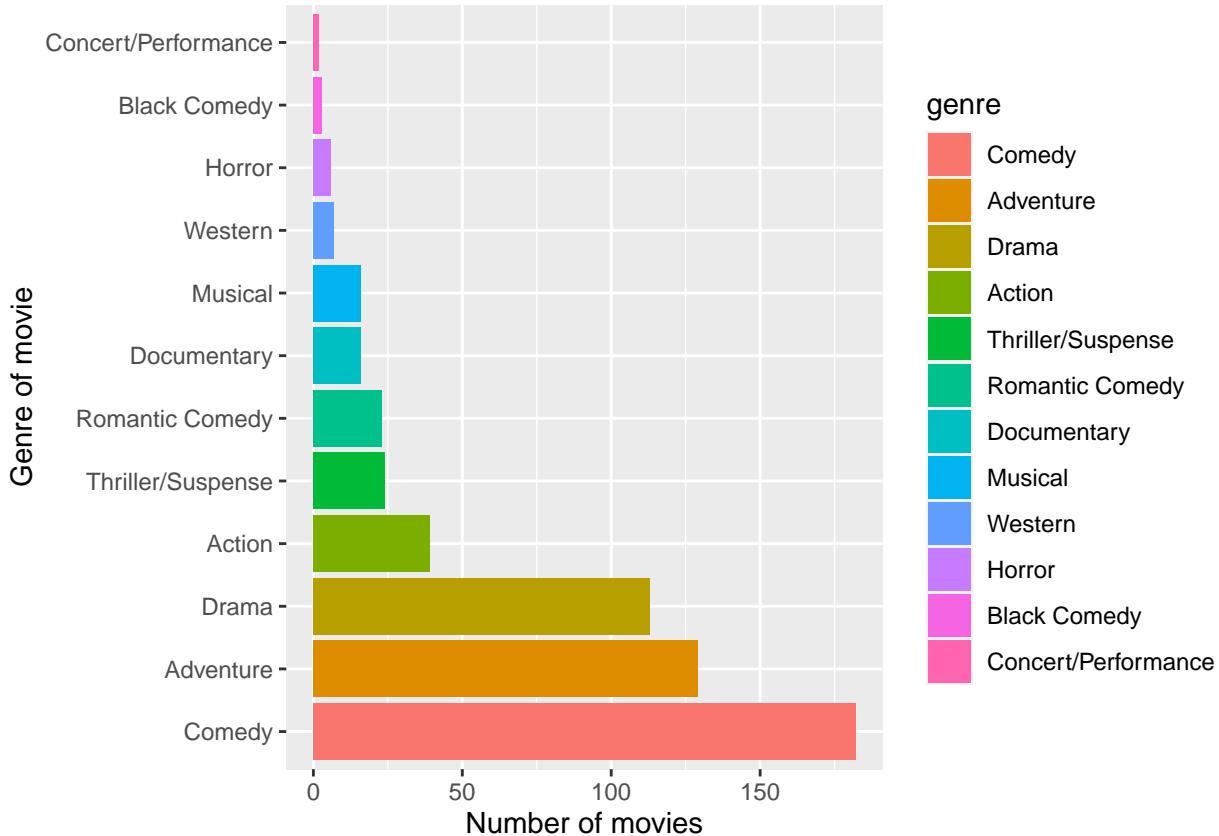
movie_title	release_month	release_year	genre	total_gross	inflation_adjusted_gross
Snow White and the Seven Dwarfs	December	1937	Musical	184925485	5228953251
Fantasia	November	1940	Musical	83320000	2187090808
Pinocchio	February	1940	Adventure	84300000	2188229052
Song of the South	November	1946	Adventure	65000000	1078510579
Cinderella	February	1950	Drama	85000000	920608730

[1] 579

Analysis

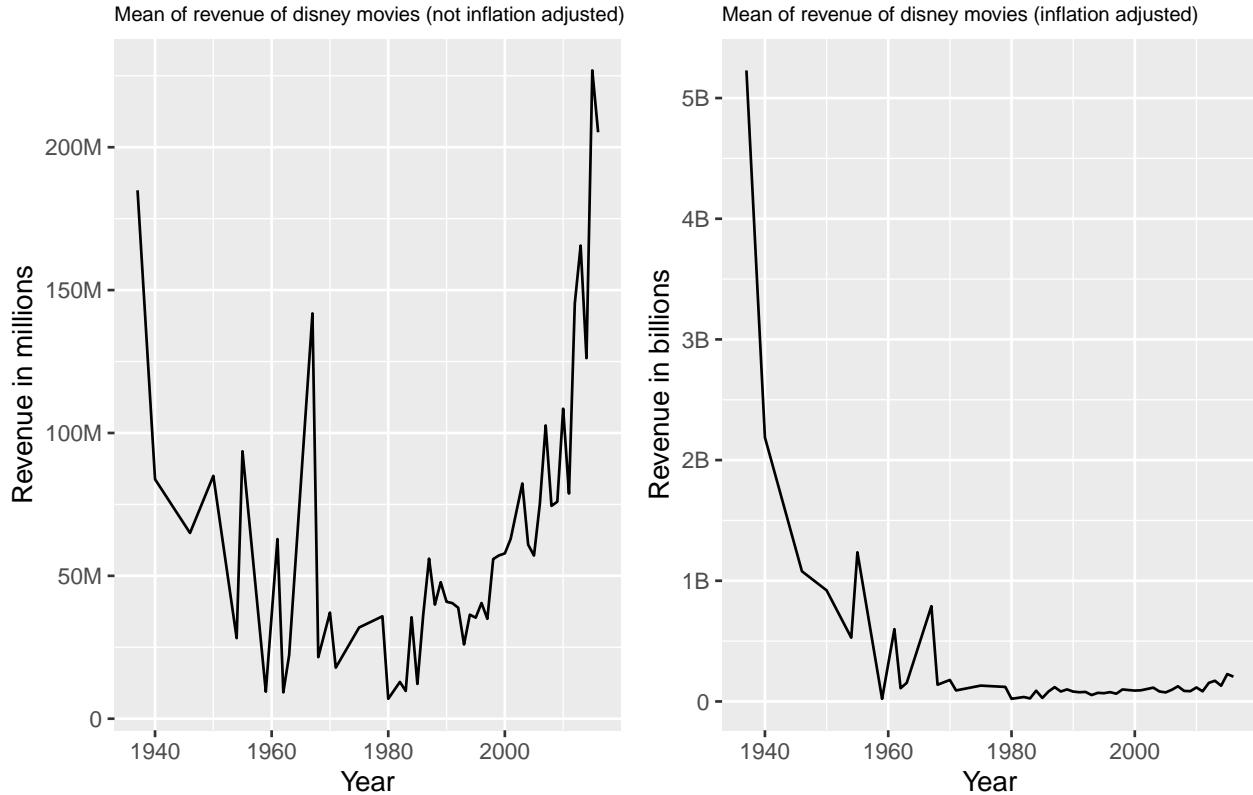
Descriptive analysis

The first thing we want to look at in this analysis is about the genre. How many movies from each genre is in our dataset ?



As we can see in this first barplot is that there are three main genre of movies that are produced by Disney. Namely **comedy**, **adventure** and **drama**. Other genre of movies are much less produces as we can see on the graph. This is an overall view of disney movies production since they stared.

The following part of this analyze concerns only the revenue of those movies, over the entire period of movies production from disney company. This start from the first movie produced, namely *Snow White and the Seven Dwarfs* in 1937, until the last *The Light Between Oceans* in 2016. Both of those graphs are showing the mean revenue of each year of movie production. The absent datas were taken off.



This first graph is about the raw/gross revenue of each movie, since it came out. In the second graph below, the amount of money generated are adjusted in term of inflation. This is important because without that you cannot really compare the first and the last movie, because of the years gap between them.

Inferential analysis

This part of this analysis will consist on seeing if the revenue of movies are significantly related to the kind of movie (comedy, adventure, drama, an so on). Because we have one categorical variable (with more than 2 groups), the genre, and a continuous (or numerical) one (revenue), then we will have to compare the difference between the means. We will proceed to an ANOVA, but we will need to check several things before knowing that. Firstly, we need to check the normality of the data because the samples are smaller than 30. As displayed in the table below, we can see that several genre don't count more than 30 observations.

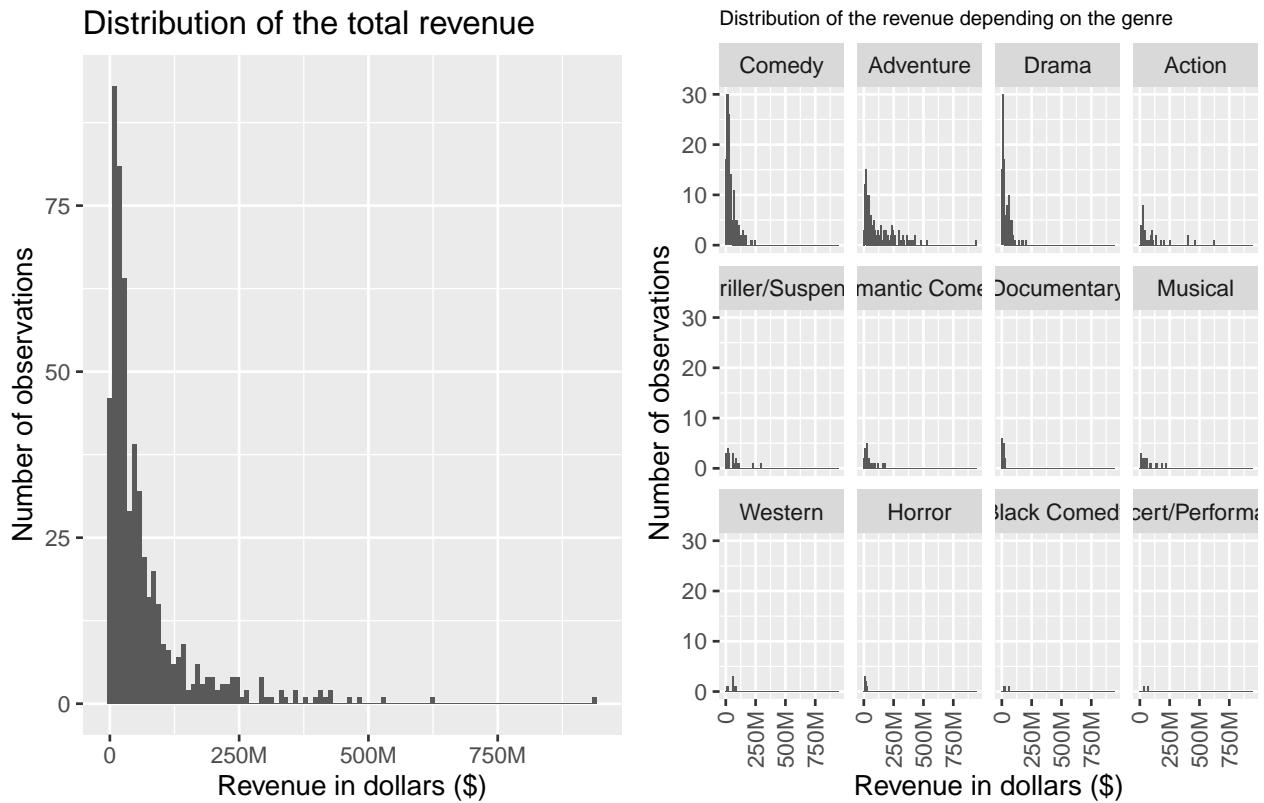
Table 2: Number of observations per genre

genre	n
Comedy	182
Adventure	129
Drama	113
Action	39
Thriller/Suspense	24
Romantic Comedy	23
Documentary	16
Musical	16
Western	7
Horror	6
Black Comedy	3
Concert/Performance	2

As we can see in the graph below, data doesn't seem that normally distributed, that's why we are computing a normality test, namely the Shapiro test. Shapiro test hypothesis are :

H0 : the data follows a normal distribution H1 : the data does not follow a normal distribution

The p-value for the shapiro test < 0.05 so we reject the null hypothesis. This means the distribution is not normal, and that's what we can observe on the histogram. As we have non-parametric data, we have to consider a non-parametric test. We will use the Kruskal-Wallis test.



```
##  
## Shapiro-Wilk normality test  
##  
## data: df_clean$total_gross  
## W = 0.63909, p-value < 2.2e-16
```

Here we will conduct the non-parametric test called Kruskal-Wallis test. In order to interpret the p-value, here are our assumptions for this test.

H0: Genres are equal in terms of revenue H1: At least one genre is different from the other genres in terms of revenue

```
kruskal.test(total_gross ~ genre, data = df_clean)
```

```
##  
## Kruskal-Wallis rank sum test  
##  
## data: total_gross by genre  
## Kruskal-Wallis chi-squared = 85.396, df = 11, p-value = 1.323e-13
```

Based on the Kruskal-Wallis test, we reject the null hypothesis and we conclude that at least one species is different in terms of flippers length (p-value $< .0001$). In order to know which group is significantly different

from the other, we conduct a post-hoc test.

```
require(FSA)

## Le chargement a nécessité le package : FSA
## ## FSA v0.9.3. See citation('FSA') if used in publication.
## ## Run fishR() for related website and fishR('IFAR') for related book.
dunnTest(total_gross ~ genre, data = df_clean, method = "holm")

## Dunn (1964) Kruskal-Wallis multiple comparison
## p-values adjusted with the Holm method.

##                                     Comparison          Z      P.unadj
## 1           Action - Adventure -0.653128350 5.136735e-01
## 2           Action - Black Comedy  0.933405845 3.506105e-01
## 3           Adventure - Black Comedy 1.161929836 2.452640e-01
## 4           Action - Comedy     3.150325928 1.630884e-03
## 5           Adventure - Comedy   5.866848071 4.441573e-09
## 6           Black Comedy - Comedy -0.005775498 9.953918e-01
## 7           Action - Concert/Performance -0.012896087 9.897107e-01
## 8           Adventure - Concert/Performance 0.154373076 8.773156e-01
## 9           Black Comedy - Concert/Performance -0.622864572 5.333735e-01
## 10          Comedy - Concert/Performance -0.795004104 4.266111e-01
## 11          Action - Documentary    4.975087608 6.521814e-07
## 12          Adventure - Documentary 6.022930110 1.712874e-09
## 13          Black Comedy - Documentary 1.458765390 1.446297e-01
## 14          Comedy - Documentary   3.532585146 4.115177e-04
## 15          Concert/Performance - Documentary 1.981841855 4.749694e-02
## 16          Action - Drama       4.090925601 4.296550e-05
## 17          Adventure - Drama    6.822855103 8.924867e-12
## 18          Black Comedy - Drama 0.342767826 7.317731e-01
## 19          Comedy - Drama      1.702215161 8.871504e-02
## 20          Concert/Performance - Drama 1.078174680 2.809558e-01
## 21          Documentary - Drama -2.685300117 7.246470e-03
## 22          Action - Horror     2.942998283 3.250502e-03
## 23          Adventure - Horror 3.376016571 7.354349e-04
## 24          Black Comedy - Horror 1.034277972 3.010062e-01
## 25          Comedy - Horror    1.770706034 7.660960e-02
## 26          Concert/Performance - Horror 1.592094760 1.113634e-01
## 27          Documentary - Horror -0.389463515 6.969333e-01
## 28          Drama - Horror     1.267078359 2.051273e-01
## 29          Action - Musical   0.443403582 6.574738e-01
## 30          Adventure - Musical 0.946954624 3.436619e-01
## 31          Black Comedy - Musical -0.679652056 4.967248e-01
## 32          Comedy - Musical   -1.626964086 1.037447e-01
## 33          Concert/Performance - Musical 0.187986559 8.508872e-01
## 34          Documentary - Musical -3.805341734 1.416085e-04
## 35          Drama - Musical   -2.351480425 1.869887e-02
## 36          Horror - Musical   -2.420967611 1.547926e-02
## 37          Action - Romantic Comedy 1.687674733 9.147368e-02
## 38          Adventure - Romantic Comedy 2.487619032 1.286014e-02
## 39          Black Comedy - Romantic Comedy -0.188232092 8.506947e-01
## 40          Comedy - Romantic Comedy -0.506938264 6.121981e-01
## 41          Concert/Performance - Romantic Comedy 0.614545191 5.388551e-01
```

```

## 42      Documentary - Romantic Comedy -3.174181560 1.502597e-03
## 43          Drama - Romantic Comedy -1.381636213 1.670834e-01
## 44          Horror - Romantic Comedy -1.847430739 6.468474e-02
## 45          Musical - Romantic Comedy  0.958579307 3.377707e-01
## 46          Action - Thriller/Suspense 1.652363373 9.846050e-02
## 47          Adventure - Thriller/Suspense 2.465264622 1.369121e-02
## 48      Black Comedy - Thriller/Suspense -0.213204449 8.311675e-01
## 49          Comedy - Thriller/Suspense -0.585720310 5.580635e-01
## 50 Concert/Performance - Thriller/Suspense  0.595171594 5.517288e-01
## 51      Documentary - Thriller/Suspense -3.248184249 1.161440e-03
## 52          Drama - Thriller/Suspense -1.472993861 1.407527e-01
## 53          Horror - Thriller/Suspense -1.888340328 5.898028e-02
## 54          Musical - Thriller/Suspense  0.920358765 3.573853e-01
## 55      Romantic Comedy - Thriller/Suspense -0.051454752 9.589632e-01
## 56          Action - Western   0.390194716 6.963926e-01
## 57          Adventure - Western  0.720257158 4.713667e-01
## 58      Black Comedy - Western  -0.578316018 5.630508e-01
## 59          Comedy - Western  -1.027390036 3.042368e-01
## 60 Concert/Performance - Western  0.211427333 8.325538e-01
## 61      Documentary - Western  -2.905932284 3.661607e-03
## 62          Drama - Western  -1.539383369 1.237107e-01
## 63          Horror - Western -2.031857098 4.216812e-02
## 64          Musical - Western  0.062955372 9.498020e-01
## 65      Romantic Comedy - Western -0.656826879 5.112922e-01
## 66      Thriller/Suspense - Western -0.625090725 5.319115e-01

## P.adj
## 1  1.000000e+00
## 2  1.000000e+00
## 3  1.000000e+00
## 4  9.132951e-02
## 5  2.842607e-07
## 6  9.953918e-01
## 7  1.000000e+00
## 8  1.000000e+00
## 9  1.000000e+00
## 10 1.000000e+00
## 11 4.108743e-05
## 12 1.113368e-07
## 13 1.000000e+00
## 14 2.469106e-02
## 15 1.000000e+00
## 16 2.663861e-03
## 17 5.890412e-10
## 18 1.000000e+00
## 19 1.000000e+00
## 20 1.000000e+00
## 21 3.840629e-01
## 22 1.787776e-01
## 23 4.339066e-02
## 24 1.000000e+00
## 25 1.000000e+00
## 26 1.000000e+00
## 27 1.000000e+00
## 28 1.000000e+00

```

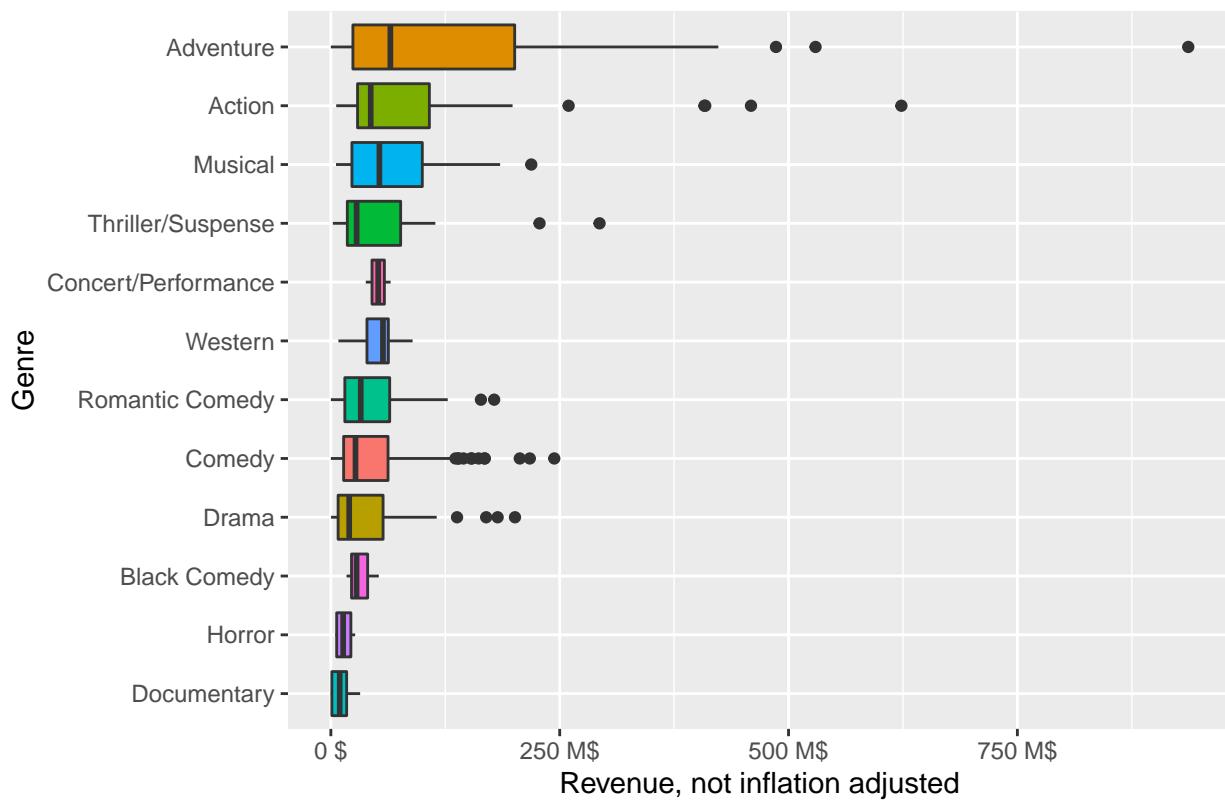
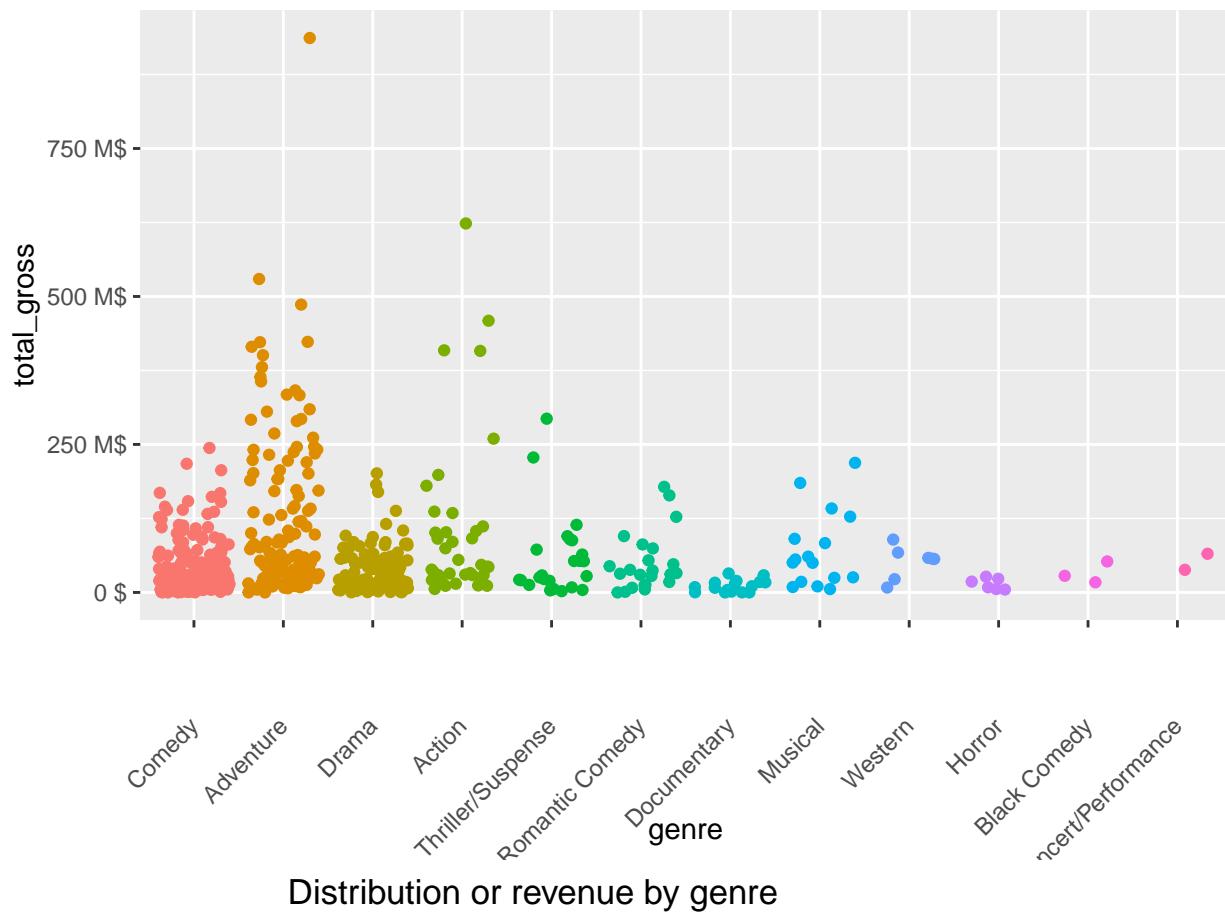
```

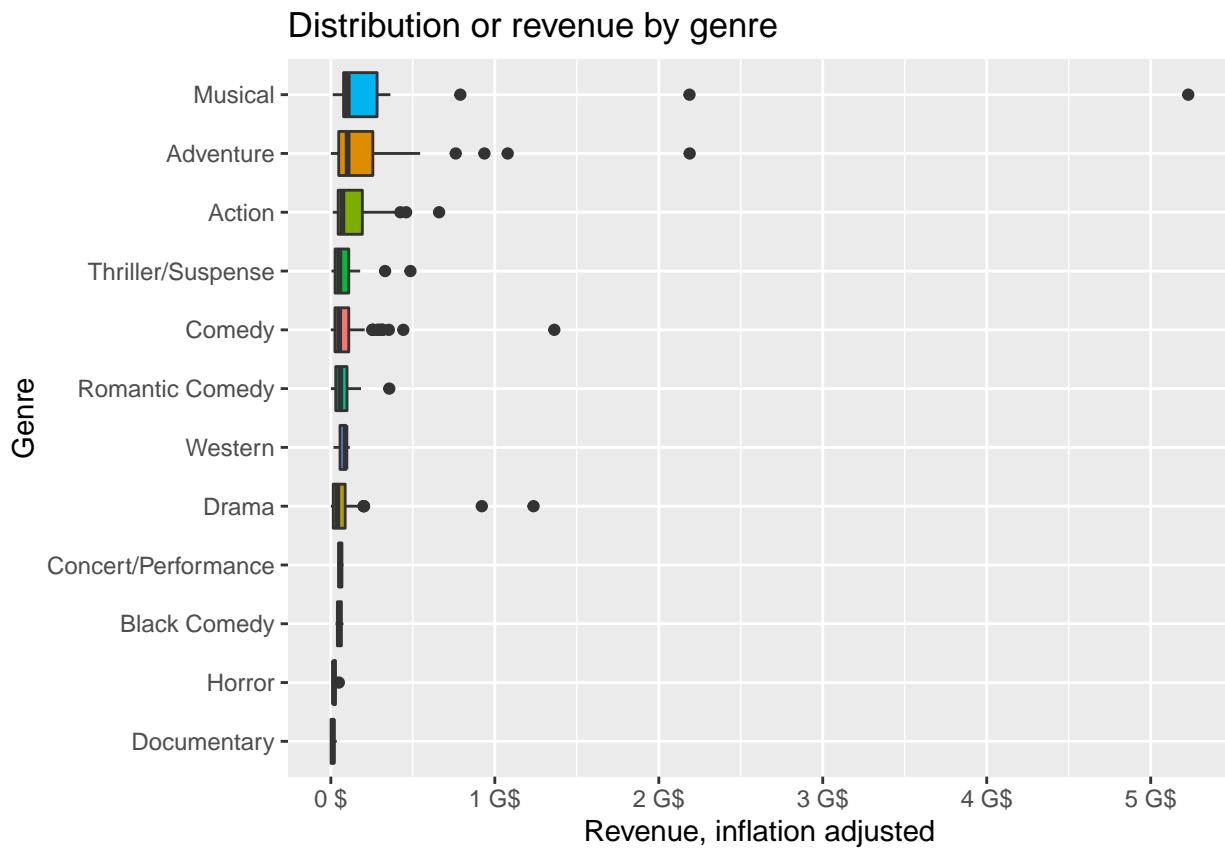
## 29 1.000000e+00
## 30 1.000000e+00
## 31 1.000000e+00
## 32 1.000000e+00
## 33 1.000000e+00
## 34 8.638121e-03
## 35 9.162448e-01
## 36 7.739628e-01
## 37 1.000000e+00
## 38 6.687273e-01
## 39 1.000000e+00
## 40 1.000000e+00
## 41 1.000000e+00
## 42 8.564801e-02
## 43 1.000000e+00
## 44 1.000000e+00
## 45 1.000000e+00
## 46 1.000000e+00
## 47 6.982517e-01
## 48 1.000000e+00
## 49 1.000000e+00
## 50 1.000000e+00
## 51 6.736354e-02
## 52 1.000000e+00
## 53 1.000000e+00
## 54 1.000000e+00
## 55 1.000000e+00
## 56 1.000000e+00
## 57 1.000000e+00
## 58 1.000000e+00
## 59 1.000000e+00
## 60 1.000000e+00
## 61 1.977268e-01
## 62 1.000000e+00
## 63 1.000000e+00
## 64 1.000000e+00
## 65 1.000000e+00
## 66 1.000000e+00

library(ggplot2)

ggplot(df_clean) +
  aes(x = genre, y = total_gross, color = genre) +
  geom_jitter() +
  scale_y_continuous(labels = label_number(scale_cut = cut_si("$"))) +
  theme(legend.position = "none") +
  theme(axis.text.x = element_text(angle = 45, vjust = 0.5, hjust=1))

```





Conclusion

References