WBO

Healy Li

12/9/2022

Introduction

Since the outbreak of Covid-19 pandemic, the movie industry and its market are severely impacted by repaid decrease of audience and the unavoidable pause of movie production. Both the movie theater and the production company, not only in US, but in worldwide, are experiencing the most difficult time in the history. In this analysis, we will investigate the effect of this pandemic on the domestic, foreign, and worldwide box office.

Method

According to pandemic timeline provided by CDC, the pandemic started around the end of 2019 and the beginning of 2020. Therefore, we decide to focus on the box office data three years before (2017-2019) and after (2010-2022) the pandemic outbreak. The basic movie information such as their title and release time are collected from IMDb. The specific box office data are obtained from BoxOfficeMojo (will be mentioned as BOM in the rest of this analysis) by using python web scraper. Movies missing both foreign and domestic box office information were excluded. Those only have foreign or domestic box office information are left because they could potentially explain how the pandemic affect a company's releasing strategy. Variables "Director" and "Actor" are deleted since they are too unique to be statistically investigated. The information of "runtime" (how long is the movie) is also deleted because movies' running time usually don't affect the final box office unless it is longer than three hours, which is not common in the movie industry. The genre, including the main genre and side genre, are kept for further investigated. The new variable "covid" was added to classify whether the movie released before or after the pandemic.

The outcome variable in this analysis is the box office of each movie. And the effect of the pandemic was investigated from the aspect of release year, genre and rating. T test was used most for analyzing the difference between movies release before and after outbreak of Covid-19. Box plot and linear regression model were used to analyze how the different factors impact the preform of each movie.

```
## # A tibble: 6 x 5
##    Year     n missing.D missing.F missing.both
##   <int> <int>    <int>     <int>        <int>
## 1  2017   180       77        74           74
## 2  2018   195       94        92           92
## 3  2019   180       84        85           82
## 4  2020   114       77        71           70
## 5  2021   148       91        86           86
## 6  2022    89       48        42           42
```

Result

The mean worldwide box office of the movies released before pandemic is significantly greater than which release after the pandemic according to the two-sample t-test (p-value = 0.0002). The same result occurs in both the comparison between mean domestic (p-value = 0.01) and foreign (p-value = 0.00086) box office. Specifically, the mean percentage of domestic box office decreased after the outbreak of Covid-19 (Table 1) while the t-test shows that there is not significant different between the two means (p-value =0.08). The mean percentage of foreign box office surprisingly increase after the pandemic and the t-test further prove that (p-value = 0.0007). Notice that although the overall box office after pandemic is lower than which before the pandemic, it was gradually increasing from 2020 to 2022 (graph 1).
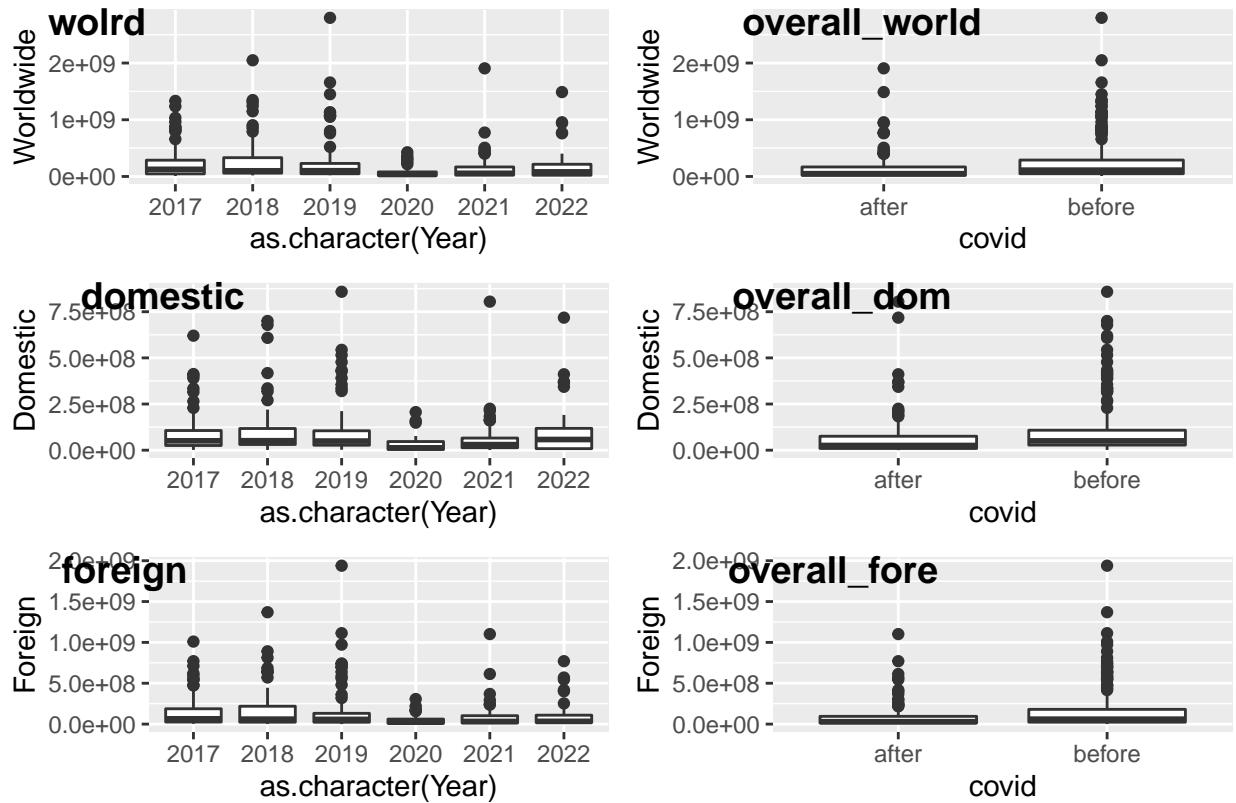

Graph 1

Table 1: Effect of Covid on Box Office

	p-value	95% CI lower bound	95% CI upper bound
Worldwide BO	0.0002038	54632747	Inf
Domestic BO	0.0101582	8351769	Inf
Foreign BO	0.0000864	38586639	Inf

```
##                         N    Msng  Mean      Std Dev   Min
##        Rank:  All       460   0     75.30     54.50     1.000
##        Rank:   Str after 153   0     65.33     52.21     1.000
##        Rank:   Str before 307   0     80.27     55.02     1.000
## Movie_Title:  All       460   0     229.5     132.6     1.000
## Movie_Title:   Str after 153   0     250.0     130.5     6.000
## Movie_Title:   Str before 307   0     219.3     132.7     1.000
```

```
##     Worldwide:  All            460     0  2.13e+08   3.232e+08   4.797e+06
##     Worldwide:  Str  after     153     0  1.452e+08  2.524e+08   4.797e+06
##     Worldwide:  Str  before    307     0  2.468e+08  3.487e+08   8.234e+06
##      Domestic:  All            460    25  8.719e+07  1.219e+08   2.036e+05
##      Domestic:  Str  after     153    18  6.752e+07  1.15e+08    2.036e+05
##      Domestic:  Str  before    307     7  9.604e+07  1.241e+08   1.193e+06
##      Dom_perc:  All            460    25    45.19      19.86      0.7000
##      Dom_perc:  Str  after     153    18    42.71      19.98      0.7000
##      Dom_perc:  Str  before    307     7    46.31      19.73      1.900
##       Foreign:  All            460     4  1.317e+08  2.125e+08   2.649e+05
##       Foreign:  Str  after     153     1  8.615e+07  1.473e+08   1.29e+06
##       Foreign:  Str  before    307     3  1.545e+08  2.354e+08   2.649e+05
##     Fore_perc:  All            460     4    57.77      21.29      0.6000
##     Fore_perc:  Str  after     153     1    62.73      22.81      9.100
##     Fore_perc:  Str  before    307     3    55.29      20.07      0.6000
##          Year:  All            460     0    2019       1.645      2017
##          Year:  Str  after     153     0    2021       0.7735     2020
##          Year:  Str  before    307     0    2018       0.8161     2017
##        Rating:  All            460     0    6.611      0.8953     2.800
##        Rating:  Str  after     153     0    6.540      0.8519     4.500
##        Rating:  Str  before    307     0    6.646      0.9155     2.800
##    main_genre:  All            460     0    3.439      2.508      1.000
##    main_genre:  Str  after     153     0    3.379      2.693      1.000
##    main_genre:  Str  before    307     0    3.469      2.415      1.000
## side_genre1:    All            460    88    5.172      3.532      1.000
## side_genre1:    Str  after     153    28    5.288      3.685      1.000
## side_genre1:    Str  before    307    60    5.113      3.459      1.000
## side_genre2:    All            460    88    9.296      5.310      1.000
## side_genre2:    Str  after     153    28    9.360      5.335      1.000
## side_genre2:    Str  before    307    60    9.263      5.308      1.000
##         covid:  All            460     0    1.667      0.4717     1.000
##         covid:  Str  after     153     0    1.000      0.0000     1.000
##         covid:  Str  before    307     0    2.000      0.0000     2.000
##                                25%       Mdn        75%        Max
##          Rank:  All            29.00     63.50      113.0      200.0
##          Rank:  Str  after     22.00     52.00      100.0      200.0
##          Rank:  Str  before    34.00     70.00      121.0      200.0
## Movie_Title:   All             114.8     229.5      344.2      458.0
## Movie_Title:   Str  after      132.0     251.0      374.0      456.0
## Movie_Title:   Str  before     110.0     210.0      331.5      458.0
##     Worldwide:  All            3.943e+07  8.701e+07  2.318e+08  2.798e+09
##     Worldwide:  Str  after     1.975e+07  4.917e+07  1.675e+08  1.907e+09
##     Worldwide:  Str  before    4.816e+07  1.112e+08  2.896e+08  2.798e+09
##      Domestic:  All            2.048e+07  4.531e+07  1.025e+08  8.584e+08
##      Domestic:  Str  after     9.141e+06  2.531e+07  7.546e+07  8.048e+08
##      Domestic:  Str  before    2.769e+07  5.025e+07  1.081e+08  8.584e+08
##      Dom_perc:  All            32.10      44.60      55.80      100.0
##      Dom_perc:  Str  after     31.65      43.10      53.70      100.0
##      Dom_perc:  Str  before    32.18      45.20      56.70      100.0
##       Foreign:  All            2.061e+07  4.675e+07  1.381e+08  1.939e+09
##       Foreign:  Str  after     1.325e+07  2.932e+07  9.506e+07  1.102e+09
##       Foreign:  Str  before    2.38e+07   6.044e+07  1.805e+08  1.939e+09
##     Fore_perc:  All            44.90      56.80      70.35      100.0
##     Fore_perc:  Str  after     47.77      59.20      79.33      100.0
```

```
##    Fore_perc:    Str  before    43.92    55.40    68.35    100.0
##          Year:  All                      2018     2019     2020     2022
##          Year:    Str  after     2020     2021     2022     2022
##          Year:    Str  before    2017     2018     2019     2019
##        Rating:  All              6.000    6.600    7.300    8.900
##        Rating:    Str  after     5.900    6.500    7.200    8.900
##        Rating:    Str  before    6.100    6.700    7.300    8.400
## main_genre:  All                 1.000    3.000    5.000    10.00
## main_genre:    Str  after        1.000    2.000    6.000    10.00
## main_genre:    Str  before       1.000    3.000    5.000    9.000
## side_genre1:  All                2.000    4.000    6.000    16.00
## side_genre1:    Str  after       2.000    4.000    6.000    15.00
## side_genre1:    Str  before      2.000    4.000    6.000    16.00
## side_genre2:  All                5.000    8.000    15.00    18.00
## side_genre2:    Str  after       5.000    8.000    15.00    17.00
## side_genre2:    Str  before      5.000    8.000    15.00    18.00
##          covid:  All             1.000    2.000    2.000    2.000
##          covid:    Str  after    1.000    1.000    1.000    1.000
##          covid:    Str  before   2.000    2.000    2.000    2.000
```

Table 2: Percnetage Difference in Domestic/Foreign Box Office

	Domestic	Foreign
p-value	0.0821916986198681	not significant
effect	0.00073896296714481	significant

Table 3: Number of Movie Didn't Release Domestically

| Year	n	missing.D	missing.F	missing.D.perc
2017	106	3	0	0.0283019
2018	103	2	0	0.0194175
2019	98	2	3	0.0204082
2020	44	7	1	0.1590909
2021	62	5	0	0.0806452
2022	47	6	0	0.1276596

```
##
##  Welch Two Sample t-test
##
## data:  BO.high$Worldwide[BO.high$covid == "before"] and BO.high$Worldwide[BO.high$covid == "after"]
## t = 2.0625, df = 13.84, p-value = 0.05846
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -18994144 944480556
## sample estimates:
##  mean of x   mean of y
## 1416078888   953335682
```

Movies in different genre were also impacted differently by the pandemic. According to graph 2, the worldwide
box office of each genre drops in various levels. Same result shows in the comparison of domestic and foreign

4

box office (graph 3 & graph 4). Action, adventure, animation, drama and horror movie seems more affected. The result of the two-sample t-test shows that the box office of comedy (p-value = 0.008), drama (p-value = 0.00005), action (p-value = 0.02) and animation (0.0006) movies before and after Covid-19 are significantly different, while horror movie previously mentioned is not as significantly different as the graph showed. Notice that fantasy and mystery movie are not discussed here since their box office data before pandemic is missing. There are no clear patterns showed when comparing box office before and after the pandemic combining main and side genres. The current information of side genre might be not sufficient to support a statistical analyze.

Graph 2
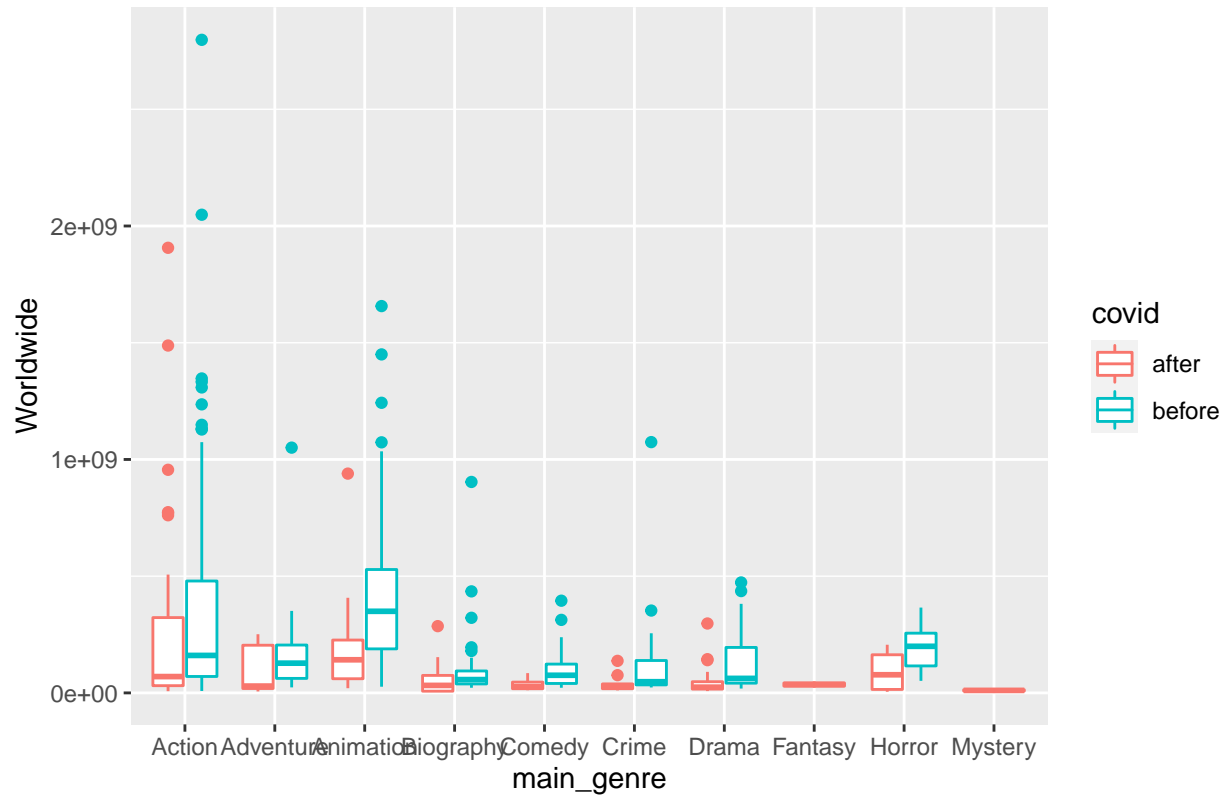
Graph 3
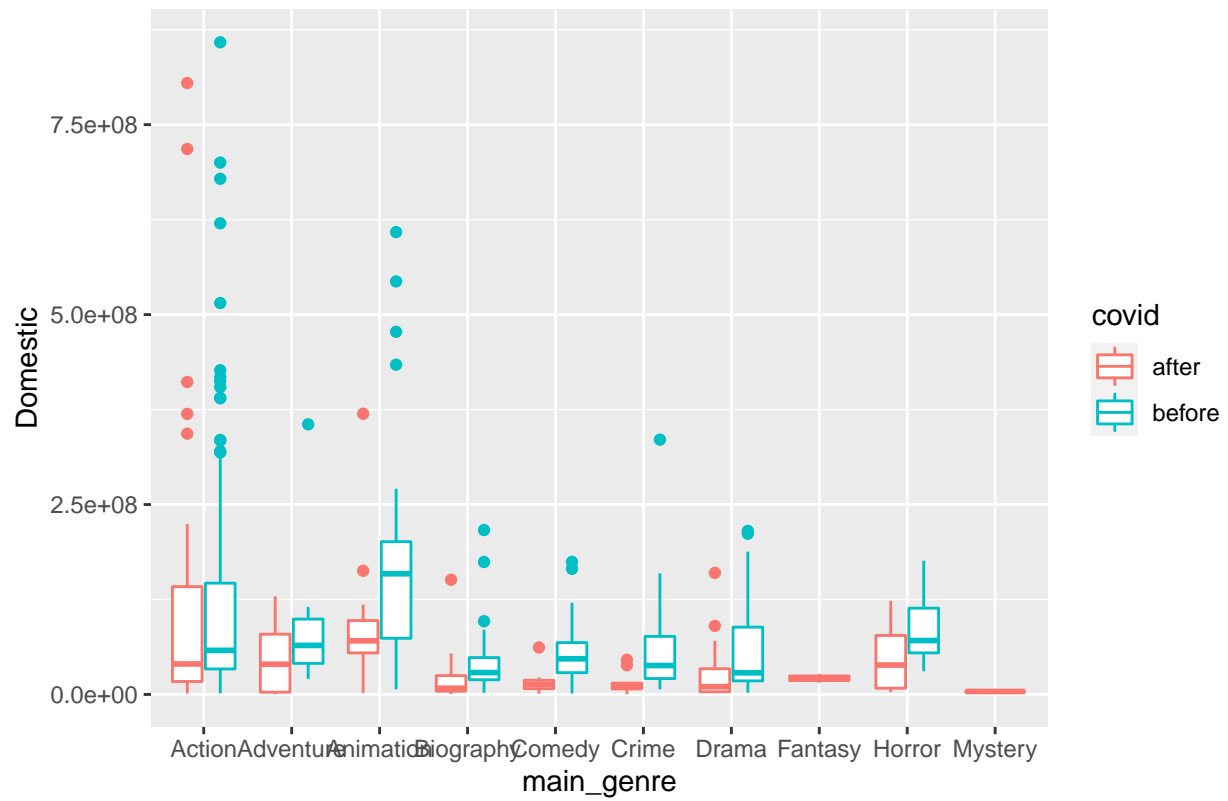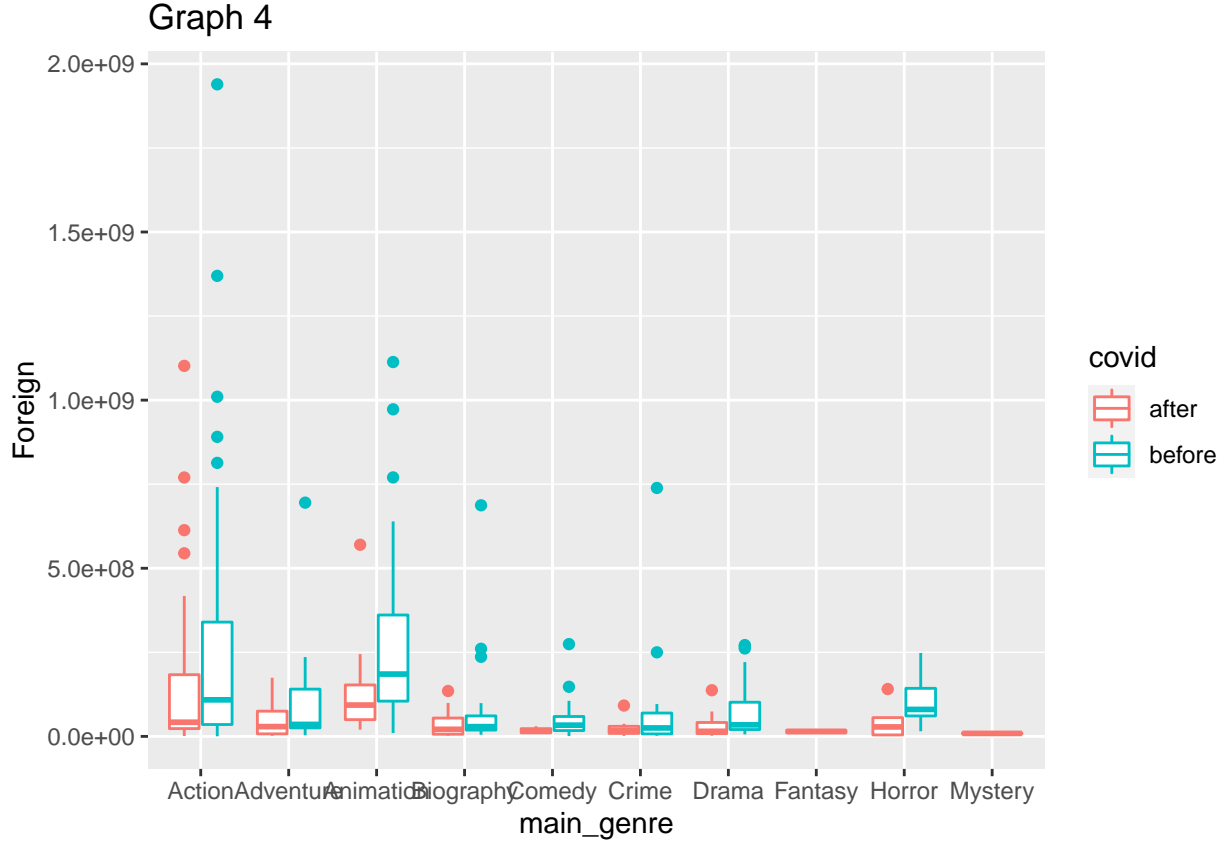
Graph 4

Table 4: Effct of Covid on Movie by Genre

	p-value	95% CI upper	95% CI lower	result
Biography	0.40821985972174	-57654594.6042637	135765934.854264	not significant
Crime	0.115811211857552	-32885580.2072072	269833998.946601	not significant
Drama	0.00815516612006213	18519058.9213138	119049730.4144	significant
Comedy	5.12427986459313e-05	33251918.6602605	86798109.8567463	significant
Adventure	0.204852023659177	-65398200.2602329	282748432.704677	not significant
Action	0.0207567503667014	20718015.516872	247825941.306657	significant
Animation	0.00617257329863006	82623874.3080143	468720539.249795	significant
Horror	0.0657493628412508	-7691649.2388569	209829870.105524	not significant

Rating of movie are positively associated with the box office regardless of the releasing time of a movie (p-value = 2.34e-07). After stratifying the box office before and after the outbreak of pandemic, similar results showed while the slope of the regression lines in after-covid group seems smaller compared to the before-covid group. The association between rating and box office of animation movie even change to negative after the outbreak (graph 5). The summary of the linear regression further insists this change. However, only the box office of action movie is significantly correlated with rating both before (p-value = 0.00257) and after (p-value = 0.00014) covid. Notice that the association between rating and box office for action movie after covid is more significant than which before covid. The box office of all the other types of movies did not show significant correlation with rating.
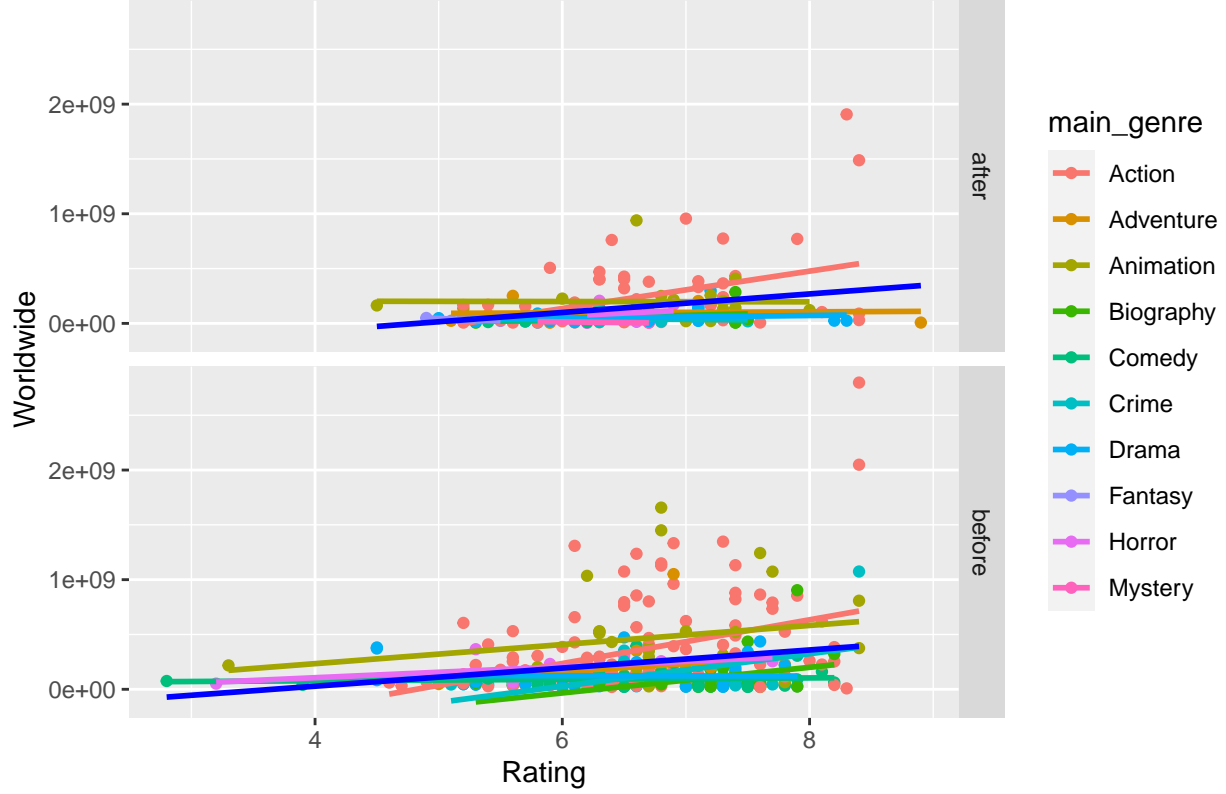
Graph 5

Table 5: Effect of Rating on Movie by Genre after Covid

	effect	p-value
Action	170509987	0.0001435
Animation	-1408056	0.9819627
Comedy	15324112	0.2944091
Drama	13090428	0.4501262

Table 6: Effect of Rating on Movie by Genre before Covid

	effect	p-value
Action	199311962	0.0000257
Animation	87137791	0.3206182
Comedy	6559229	0.5288247
Drama	-1554785	0.9406083

Discussion

It is certain that the box office of movie was significantly lowered by the pandemic, not only in domestic area, but also in foreign country. This might because of the shutdown of movie theater and the self-quarantine/physical distancing police in worldwide. In order to avoid the transmission of disease, people are less likely to attend gathering activities. There was no significant difference in the percentage of domestic

box office after pandemic, while the percentage of foreign box office of movie increased, which is contradict with the trend of worldwide box office. This might because of, firstly, the larger size of the foreign market and, secondly, the change of releasing strategy after the pandemic. The percentage of the movie missing domestic box office are higher after covid, meaning that some movies did not release in domestic area (U.S). The animation movie Turning Red is one of the examples. Instead of releasing in both domestic and foreign area, Disney choose to release it directly on their stream platform Disney+, which, in a certain level, lower the box office of this movie.

The effect of the pandemic on movies are also varies with their genre. Biography, horror and crime movie were not significantly impact. The relatively stable fan group of such kind of movie could be one of the reasons. While other type of film, including action, comedy, drama and animation, are highly relay on the teenagers, kids and family audience, which usually go to cinema with a more-than-four-people group. This way of social activity become less popular under the pandemic, which gives a huge impact on box office. Among those four categories of movie, animation is the most affected type with the smallest p-value (0.006). This is reasonable because young kids are one of the most vulnerable groups during pandemic, meanwhile they are also the major audience of animation movie. Parents tends to keep them at home rather than bring them out, which increases their probability of getting infected. Because of the loss of the major target audience, animation company like Disney, started change their strategy of releasing new animation movies, which further lower the box office of animation film. The influence of pandemic on adventure film is not significant, which is kind of surprise since this is also a kind of movie that depend on family audiences largely. A possible explanation is the small sample size of adventure film. Our dataset shows that only 22 adventure movies were released during 2017-2022, which might potentially cause inaccurate result in this analysis.

Rating does facilitate the box office of a movie both before and after the covid, while its effect is weaker to the box office of the movie released after the pandemic. Higher rating did bring some positive effects to the movie industry, but still limit. The influence of rating on different type of movie after covid also varies. We only focus on action, animation, comedy and drama movie in this case since the last part shows that there was no significant difference between the performance of the other four types of movies before and after covid. According to the two-sample t-test, action movie is most affected by rating. This genre of movie, such as The Avengers, Star Wars, usually come with large amount of visual effect, which means they usually requires a better movie watching environment with a high quality of projection and sound effect. Therefore, with the push a high rating, audience of action movie will be more likely to step into the movie theater even under the threat of Covid-19.

Overall, the negative influence of Covid-19 on movie industry is undoubted. Both domestic and foreign box office of movie were severely lowered after the outbreak. The lifestyle and the releasing strategy of movie company might be two of the many reasons. Action, animation, drama and comedy, which is most of the movie release in each year, are the most affected genres during this pandemic. Higher rating of movies helps the performance of movie. However, under such severe pandemic, its assistance is still limited. Fortunately, the movie industry is recovering as covid mutating to a less harmful version, and the top blockbuster such as Spiderman: No Way Home, Tenet, Doctor Strange, still shows powerful attraction to the audiences. Their performance is not significantly different with the blockbuster released before pandemic.