電影票房預測

陳昱鈞
張頌平
曹書誠

取得電影排名

Rank	Movie Title (click to view)	Studio Artsploitati	Total Gross / Theaters		Opening / Theaters		Open	Close
1	**Star Wars: The Last Jedi**	BV	**$620,181,382**	4,232	$220,009,584	4,232	12/15	4/19
2	**Beauty and the Beast (2017)**	BV	**$504,014,165**	4,210	$174,750,616	4,210	3/17	7/13
3	**Wonder Woman**	WB	**$412,563,408**	4,165	$103,251,471	4,165	6/2	11/9
4	**Jumanji: Welcome to the Jungle**	Sony	**$404,515,480**	3,849	$36,169,328	3,765	12/20	-
5	**Guardians of the Galaxy Vol. 2**	BV	**$389,813,101**	4,347	$146,510,104	4,347	5/5	9/21

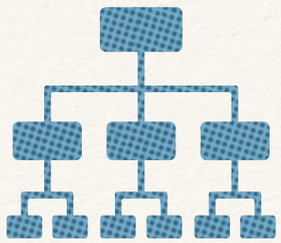
- 從BoxOfficeMojo網站抓取美國2013至2017共5年票房前100名的電影

取得各個電影資訊

Star Wars: The Last Jedi

Domestic Total Gross: **$620,181,382**	
Distributor: **Buena Vista**	Release Date: **December 15, 2017**
Genre: **Sci-Fi Fantasy**	Runtime: **2 hrs. 31 min.**
MPAA Rating: **PG-13**	Production Budget: **N/A**

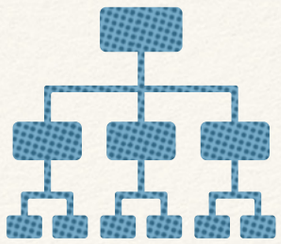
- Genre：電影類型
- Runtime：電影片長
- MPAA Rating：電影分級



新變量

- $Playing_day：確切總上映天數（下檔日期−上映日期）
- $Running.time：片長（以分鐘計）
- $Running.length：依片長分等級

< 90	90 ~ 120	120 ~ 150	> 150
short	medium	medium long	long

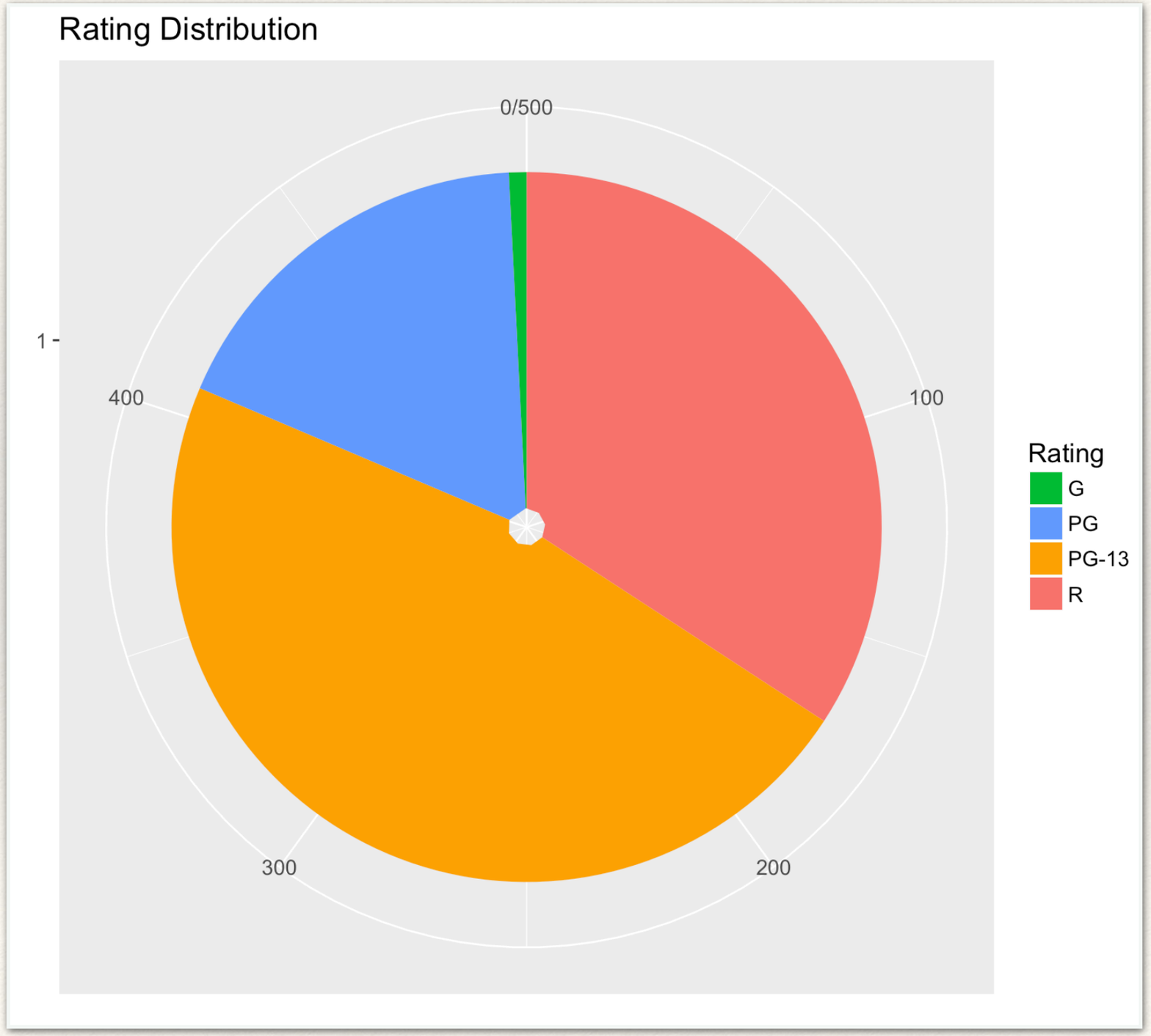


新變量

- $Gross_level：票房收入（以億計，取小數點前一位）
- $Ranking：將每年排名前100名再分級

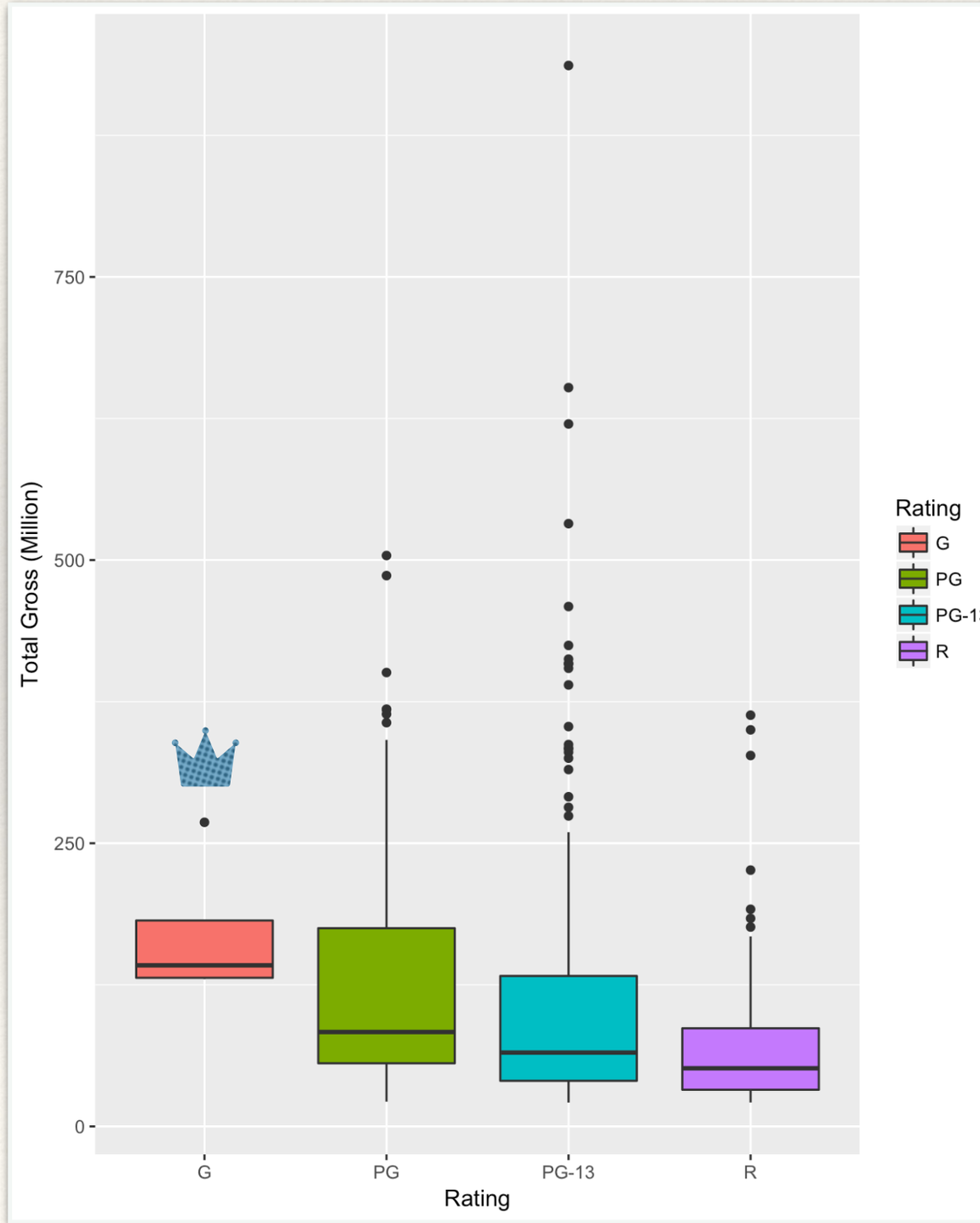
Yearly_Ranking	1 - 10	11 - 20	21 - 30	31 - 40	40 -
Ranking	1	2	3	4	5

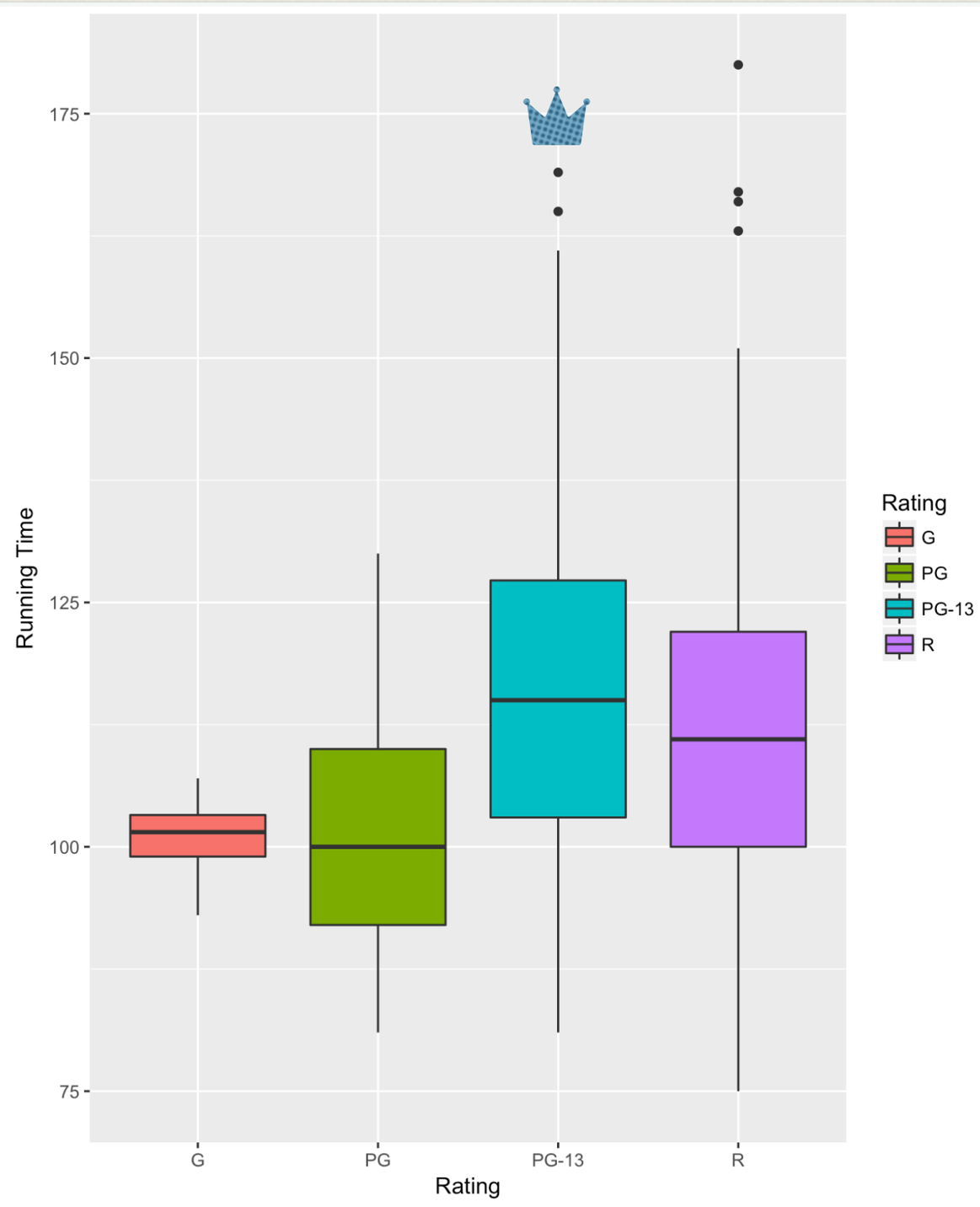
電影分級



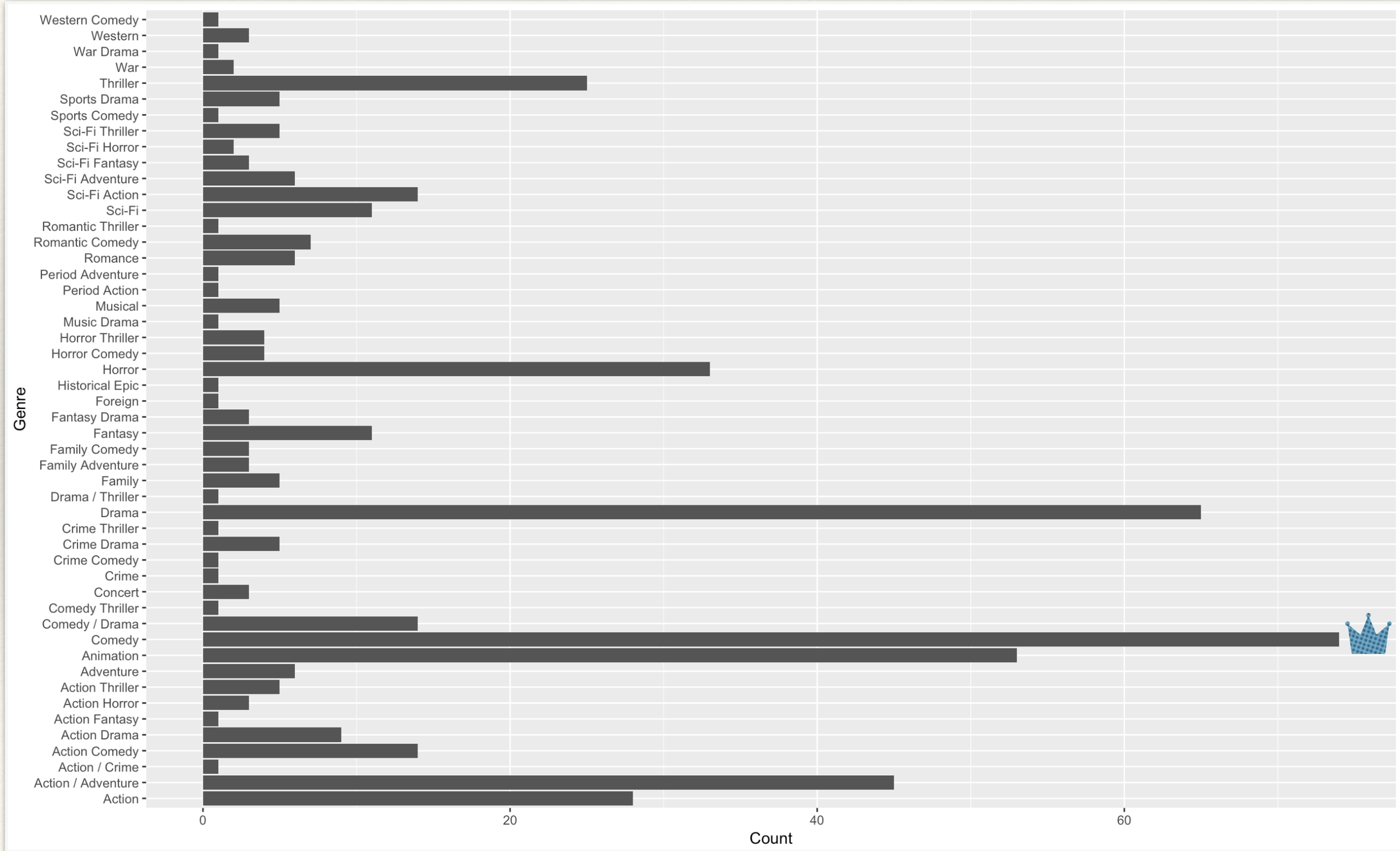
- General Audiences （一般觀眾）
- Parental Guidance Suggested （建議家長陪同）
- Parents Strongly Cautioned （家長須特別注意）
- Restricted （限制級）

各電影分級的總票房 / 片長

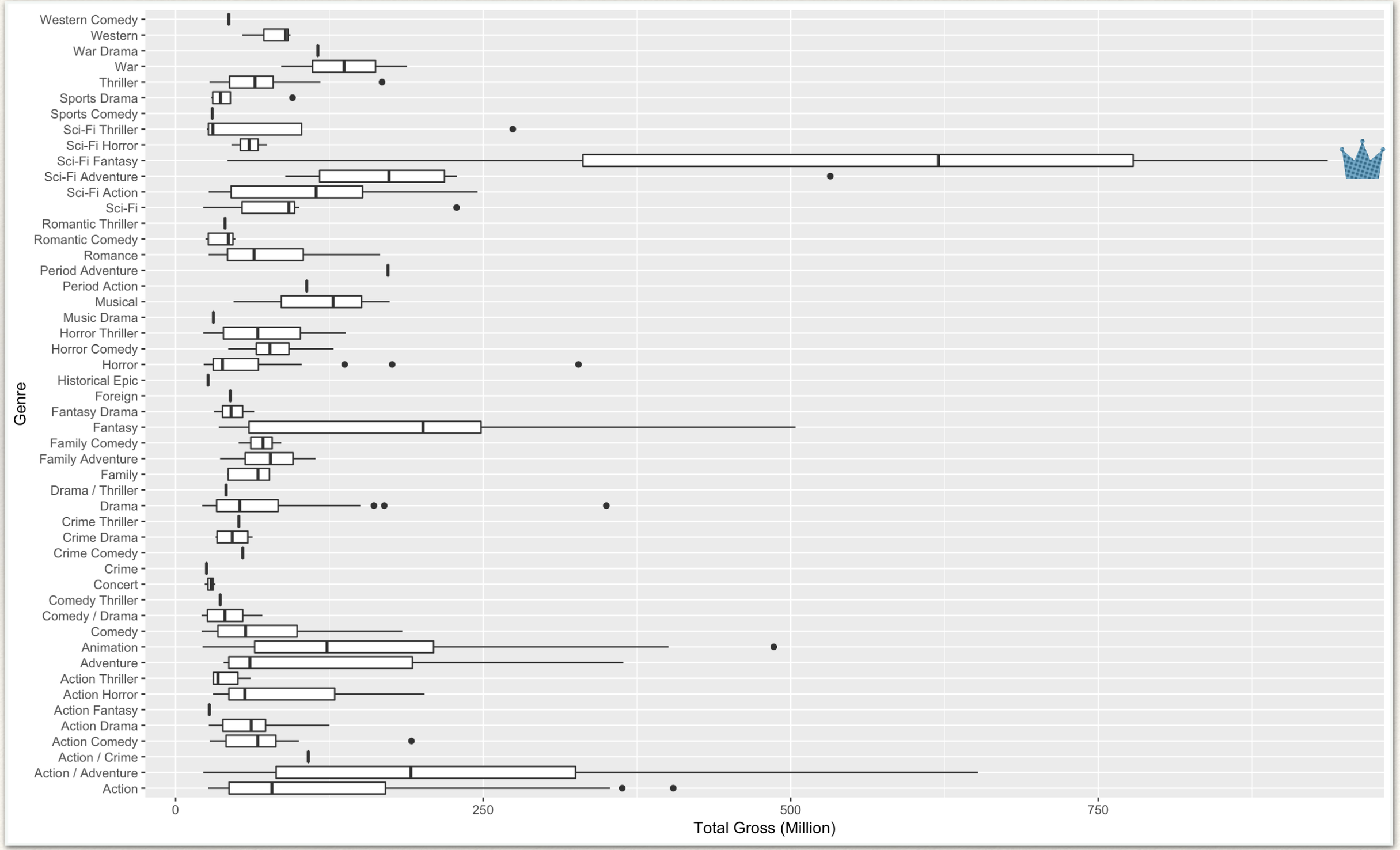




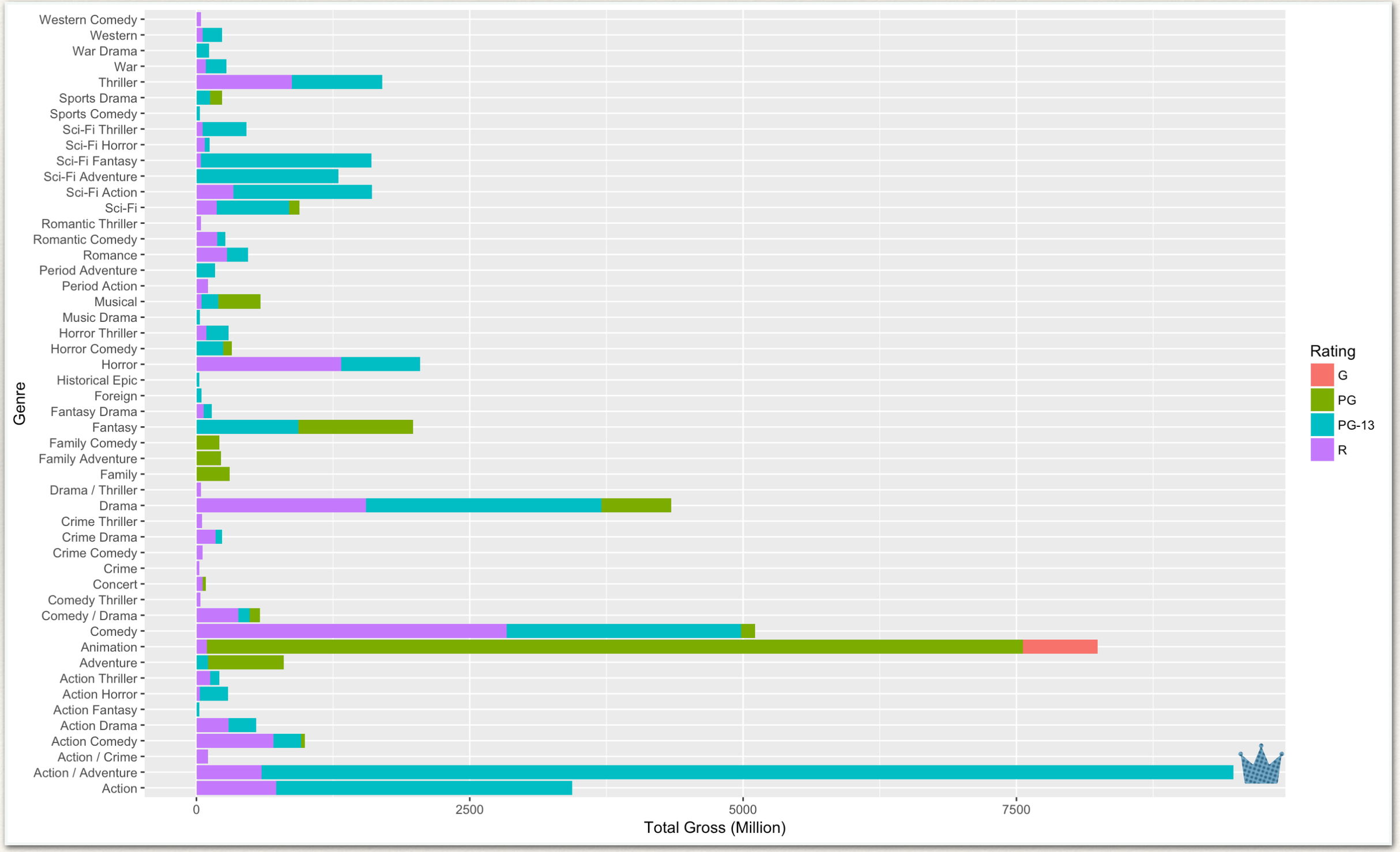
電影類型數量


Western Comedy
Western
War Drama
War
Thriller
Sports Drama
Sports Comedy
Sci-Fi Thriller
Sci-Fi Horror
Sci-Fi Fantasy
Sci-Fi Adventure
Sci-Fi Action
Sci-Fi
Romantic Thriller
Romantic Comedy
Romance
Period Adventure
Period Action
Musical
Music Drama
Horror Thriller
Horror Comedy
Horror
Historical Epic
Foreign
Fantasy Drama
Fantasy
Family Comedy
Family Adventure
Family
Drama / Thriller
Drama
Crime Thriller
Crime Drama
Crime Comedy
Crime
Concert
Comedy Thriller
Comedy / Drama
Comedy
Animation
Adventure
Action Thriller
Action Horror
Action Fantasy
Action Drama
Action Comedy
Action / Crime
Action / Adventure
Action
Genre
0
20
40
60
Count

電影類型的總票房


Western Comedy
Western
War Drama
War
Thriller
Sports Drama
Sports Comedy
Sci-Fi Thriller
Sci-Fi Horror
Sci-Fi Fantasy
Sci-Fi Adventure
Sci-Fi Action
Sci-Fi
Romantic Thriller
Romantic Comedy
Romance
Period Adventure
Period Action
Musical
Music Drama
Horror Thriller
Horror Comedy
Horror
Historical Epic
Foreign
Fantasy Drama
Fantasy
Family Comedy
Family Adventure
Family
Drama / Thriller
Drama
Crime Thriller
Crime Drama
Crime Comedy
Crime
Concert
Comedy Thriller
Comedy / Drama
Comedy
Animation
Adventure
Action Thriller
Action Horror
Action Fantasy
Action Drama
Action Comedy
Action / Crime
Action / Adventure
Action
Genre
0
250
500
750
Total Gross (Million)

各電影類型分級的總票房



新變量

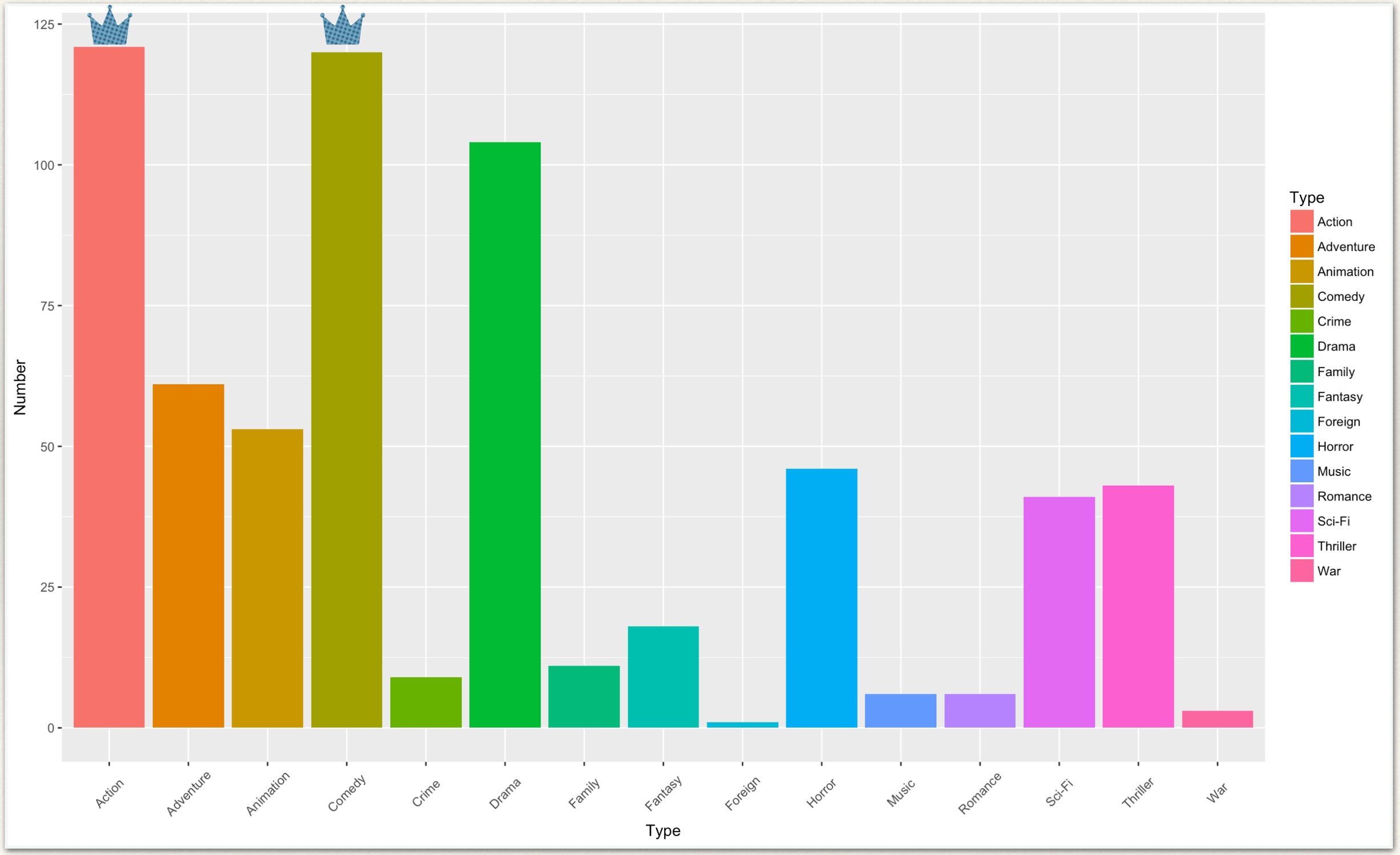
- Genre:原本的電影類型

Genre
- Action
- Action / Adventure
- Action / Crime
- Action Comedy
- Action Drama
- Action Fantasy
- Action Horror
- Action Thriller
- Adventure
- Animation
- Comedy
- Comedy / Drama
- Comedy Thriller
- Concert
- Crime
- Crime Comedy
- Crime Drama
- Crime Thriller
- Drama
- Drama / Thriller
- Family
- Family Adventure
- Family Comedy
- Fantasy
- Fantasy Drama
- Foreign
- Historical Epic
- Horror
- Horror Comedy
- Horror Thriller
- Music Drama
- Musical
- Period Action
- Period Adventure
- Romance
- Romantic Comedy
- Romantic Thriller
- Sci-Fi
- Sci-Fi Action
- Sci-Fi Adventure
- Sci-Fi Fantasy
- Sci-Fi Horror
- Sci-Fi Thriller
- Sports Comedy
- Sports Drama
- Thriller
- War
- War Drama
- Western
- Western Comedy

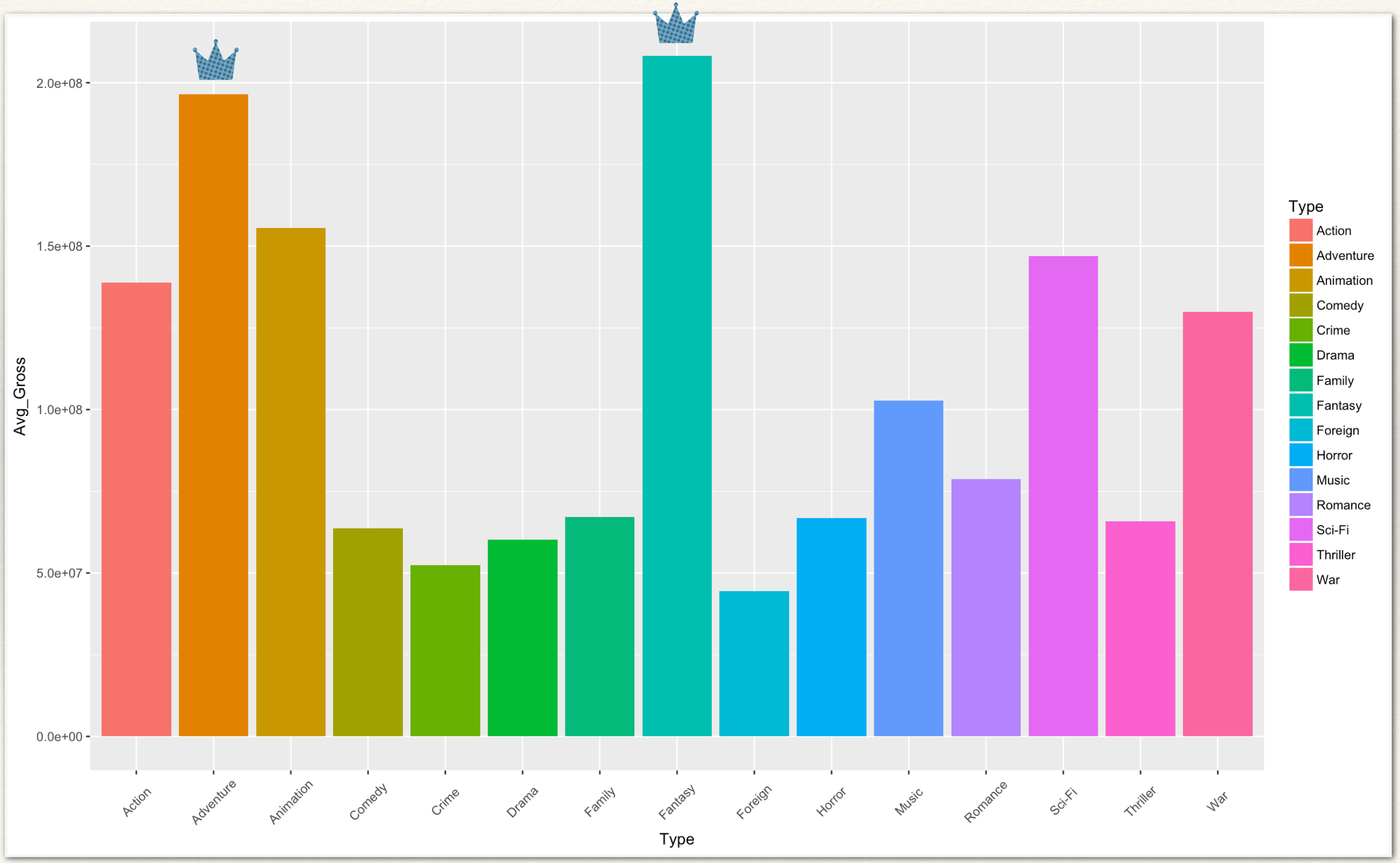
Type:簡化後的電影類型

Type
- Action
- Adventure
- Animation
- Comedy
- Crime
- Drama
- Family
- Fantasy
- Foreign
- Horror
- Music
- Romance
- Sci-Fi
- Thriller
- War

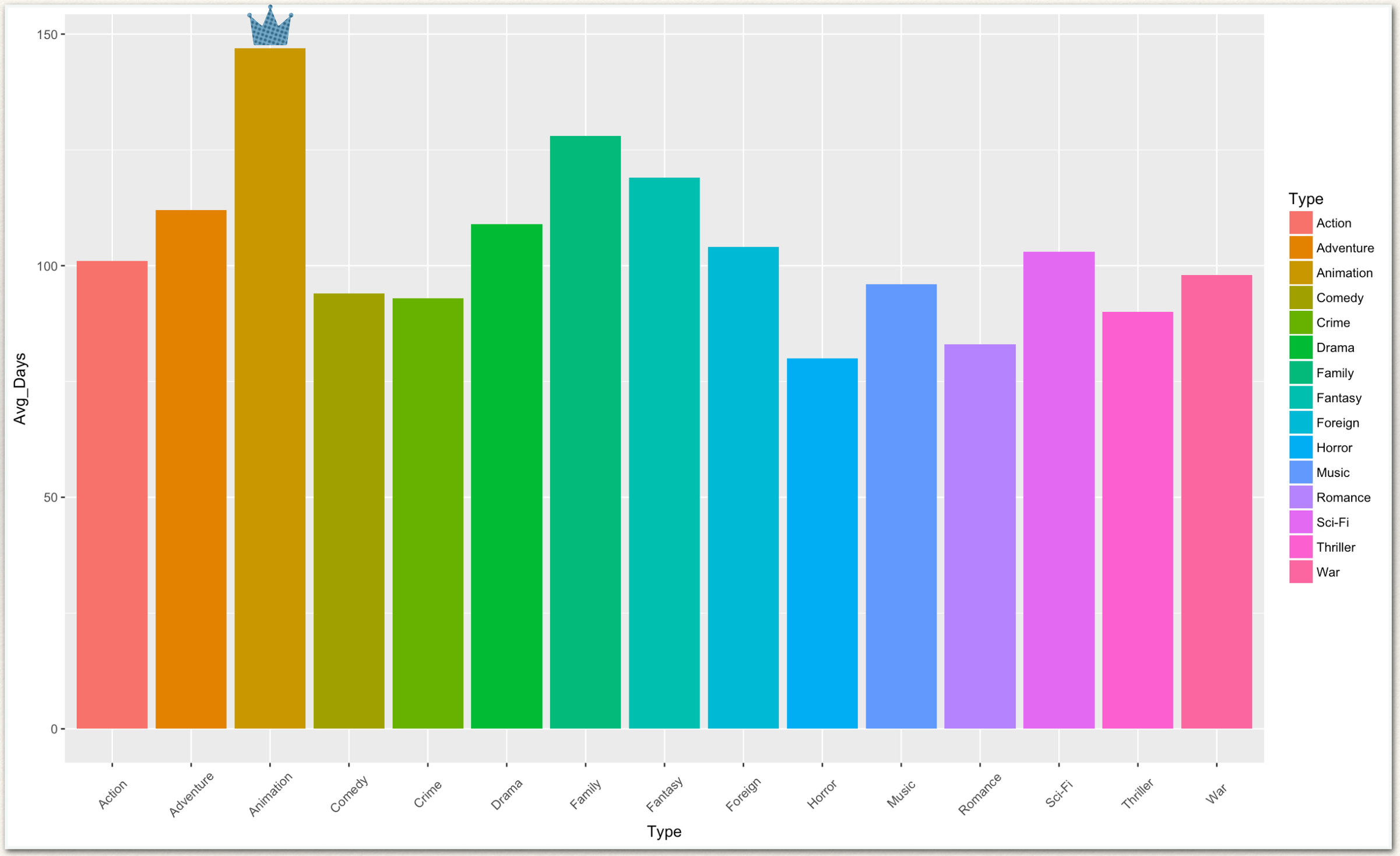
電影類型數量


125
100
75
50
25
0
Number
Action
Adventure
Animation
Comedy
Crime
Drama
Family
Fantasy
Foreign
Horror
Music
Romance
Sci-Fi
Thriller
War
Type
Type
Action
Adventure
Animation
Comedy
Crime
Drama
Family
Fantasy
Foreign
Horror
Music
Romance
Sci-Fi
Thriller
War

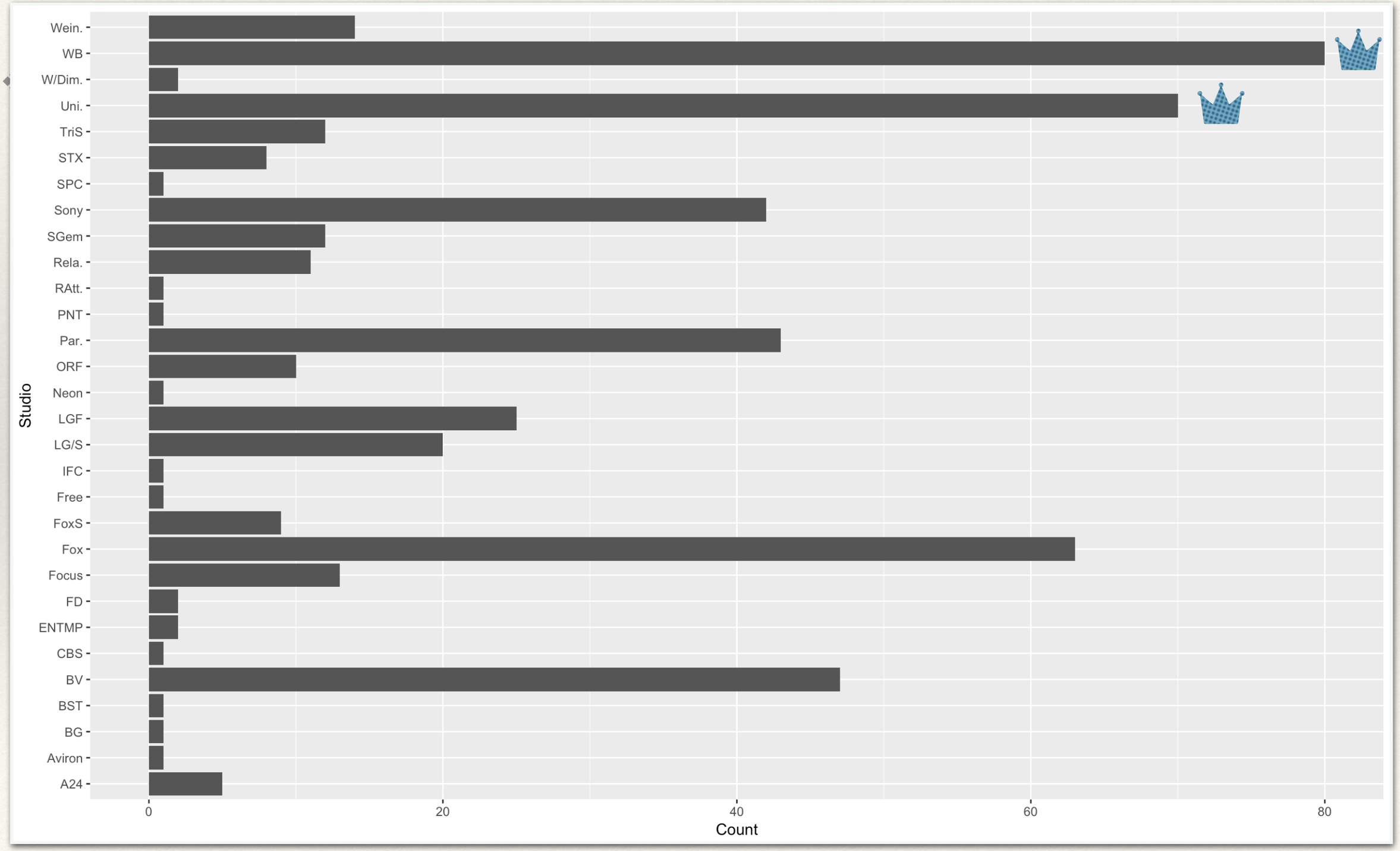
各電影類型的平均票房


2.0e+08
1.5e+08
1.0e+08
5.0e+07
0.0e+00
Avg_Gross
Action
Adventure
Animation
Comedy
Crime
Drama
Family
Fantasy
Foreign
Horror
Music
Romance
Sci-Fi
Thriller
War
Type
Type
Action
Adventure
Animation
Comedy
Crime
Drama
Family
Fantasy
Foreign
Horror
Music
Romance
Sci-Fi
Thriller
War

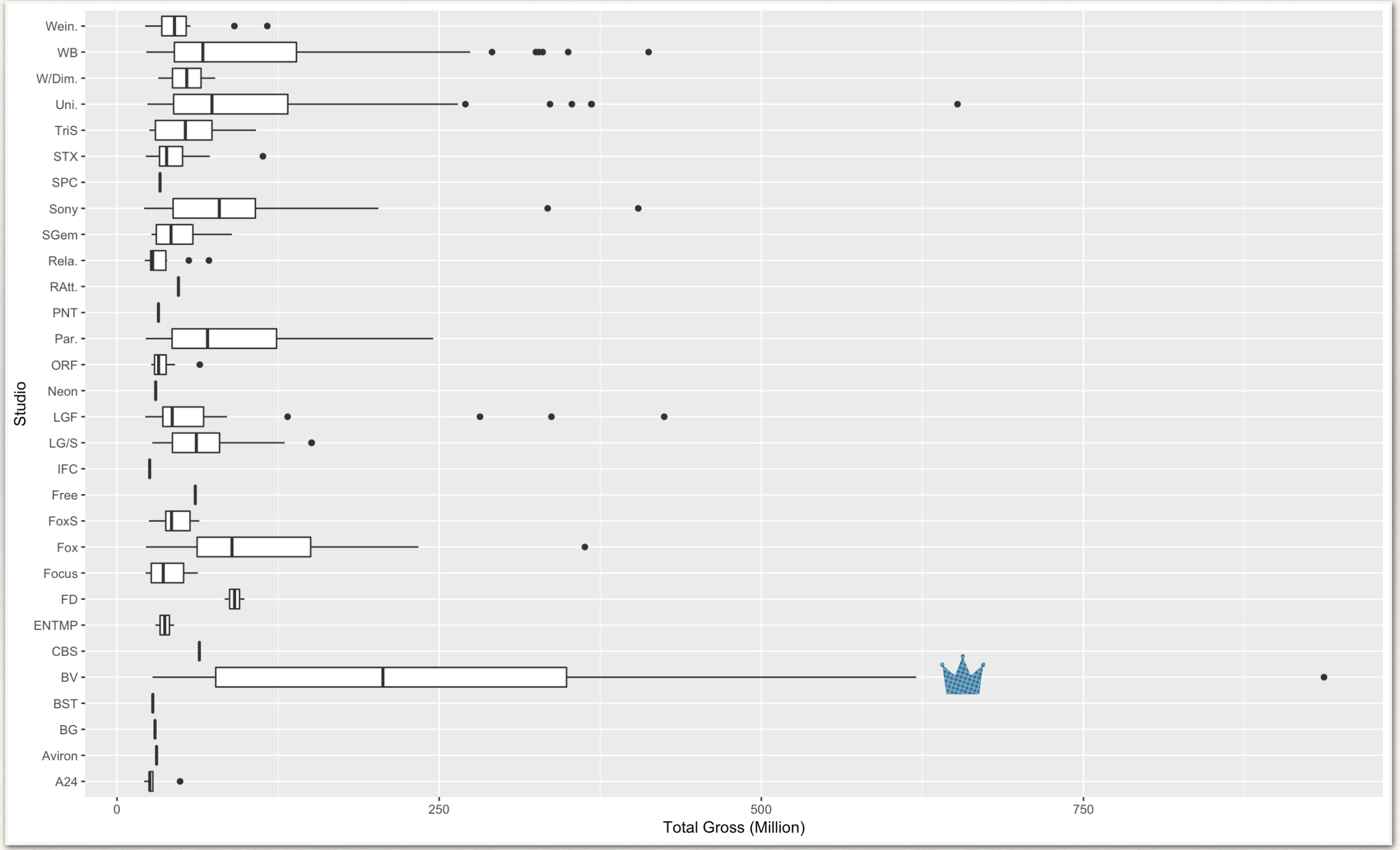
各電影類型的平均上映天數



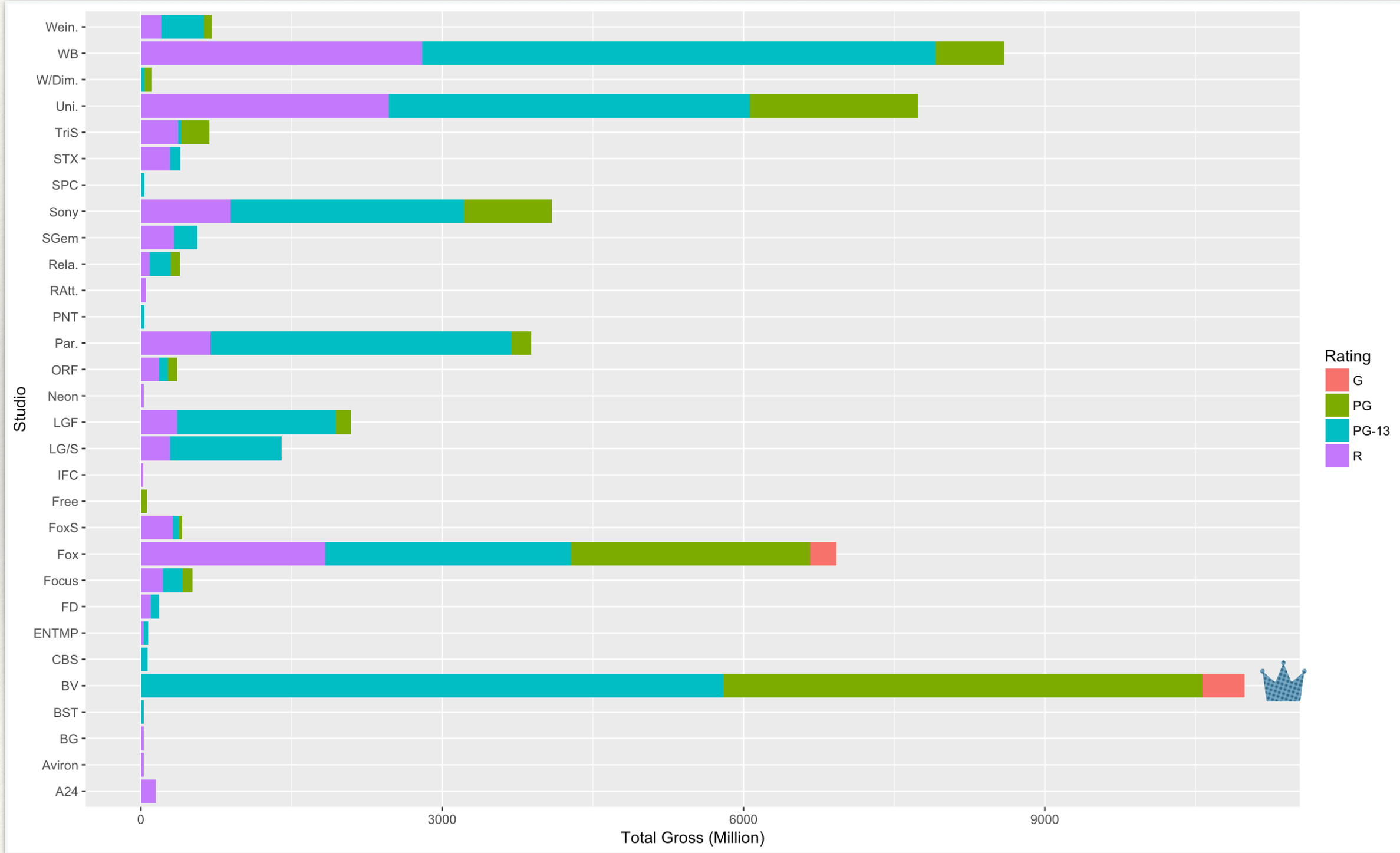
各製片公司的電影數量


Wein.
WB
W/Dim.
Uni.
TriS
STX
SPC
Sony
SGem
Rela.
RAtt.
PNT
Par.
ORF
Neon
LGF
LG/S
IFC
Free
FoxS
Fox
Focus
FD
ENTMP
CBS
BV
BST
BG
Aviron
A24
Studio
0
20
40
60
80
Count

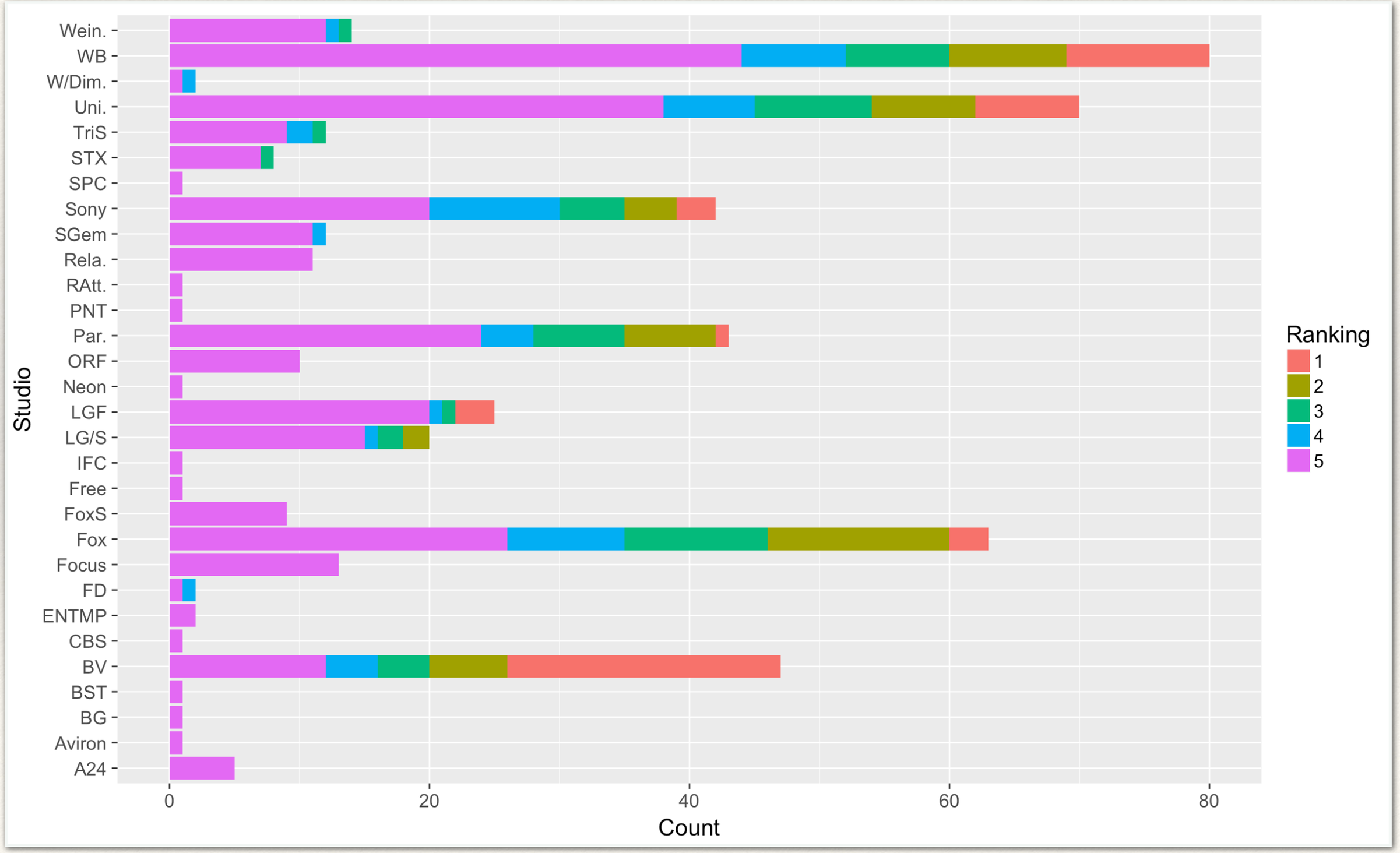
各製片公司的平均票房


Wein.
WB
W/Dim.
Uni.
TriS
STX
SPC
Sony
SGem
Rela.
RAtt.
PNT
Par.
ORF
Neon
LGF
LG/S
IFC
Free
FoxS
Fox
Focus
FD
ENTMP
CBS
BV
BST
BG
Aviron
A24
Studio
0
250
500
750
Total Gross (Million)

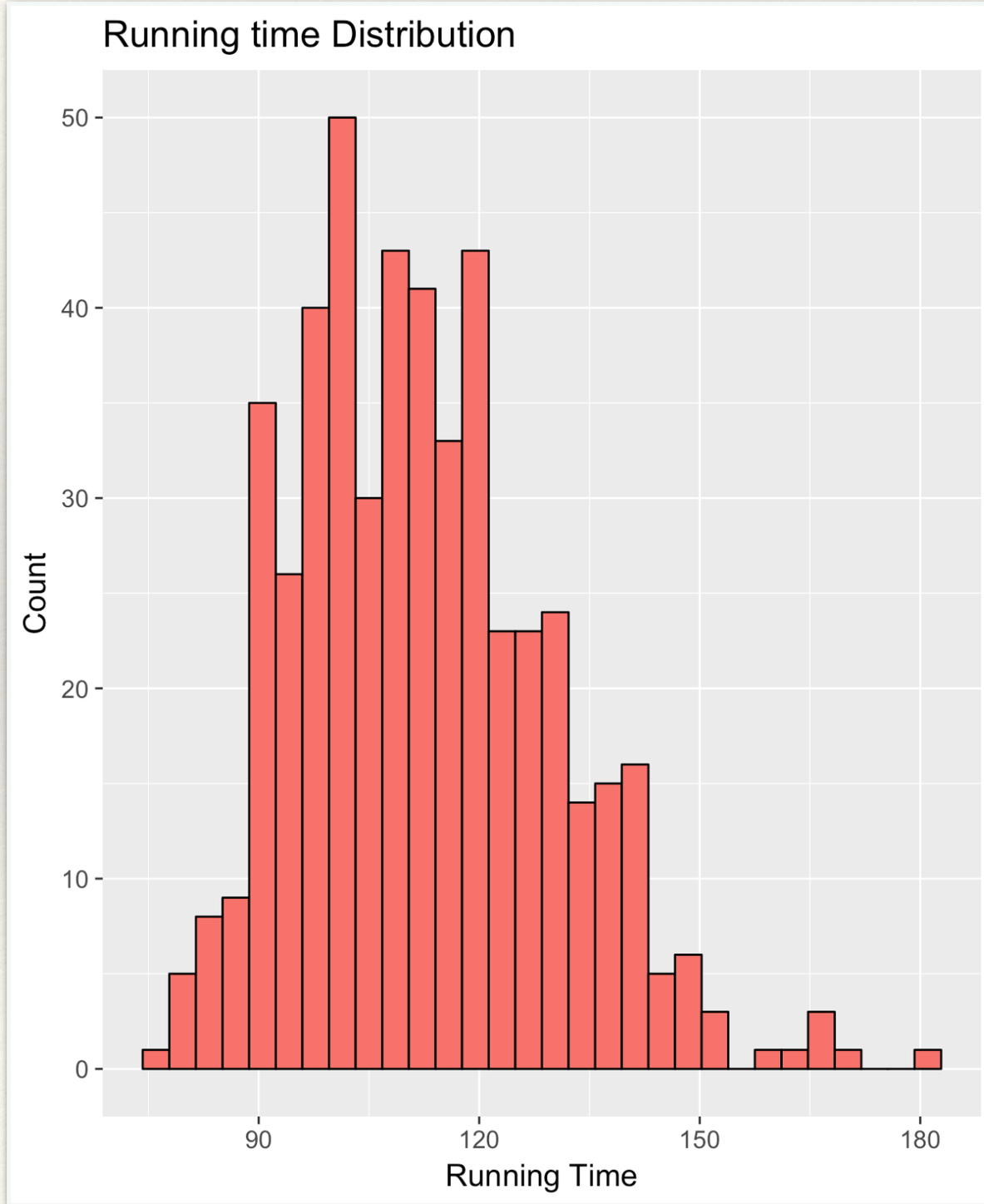
各製片公司分級的總票房

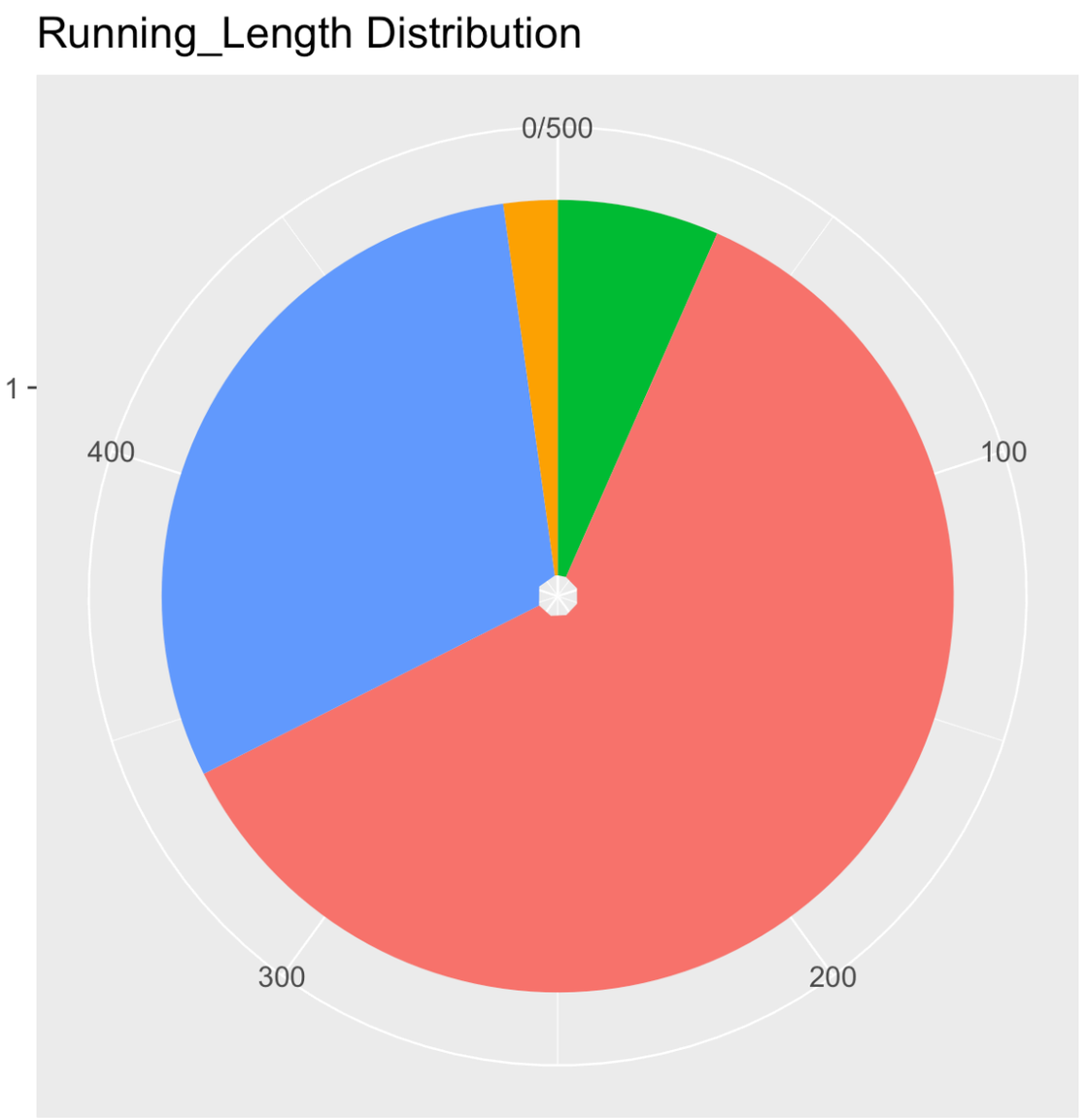


各製片公司的電影排名數量

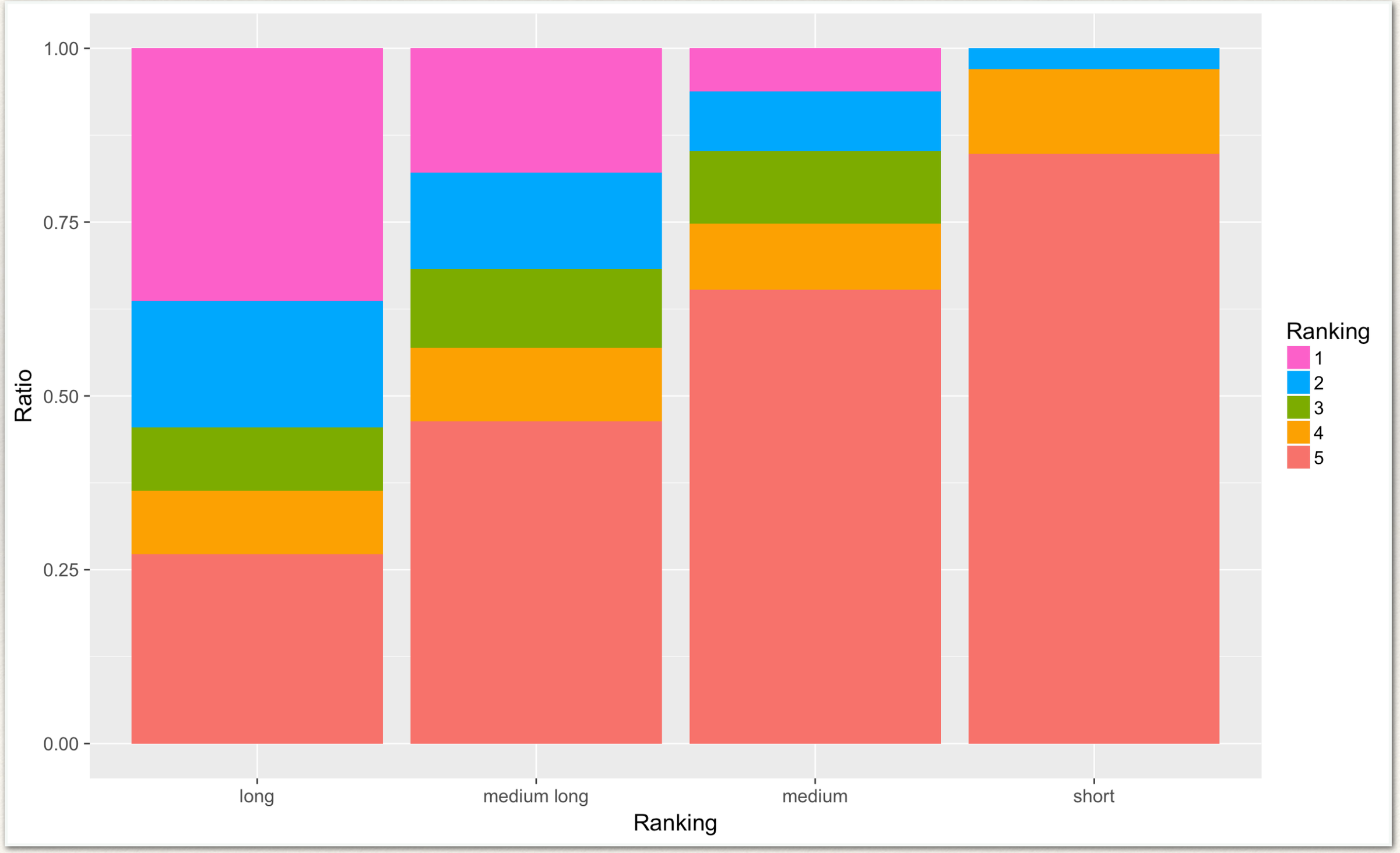


電影片長分佈

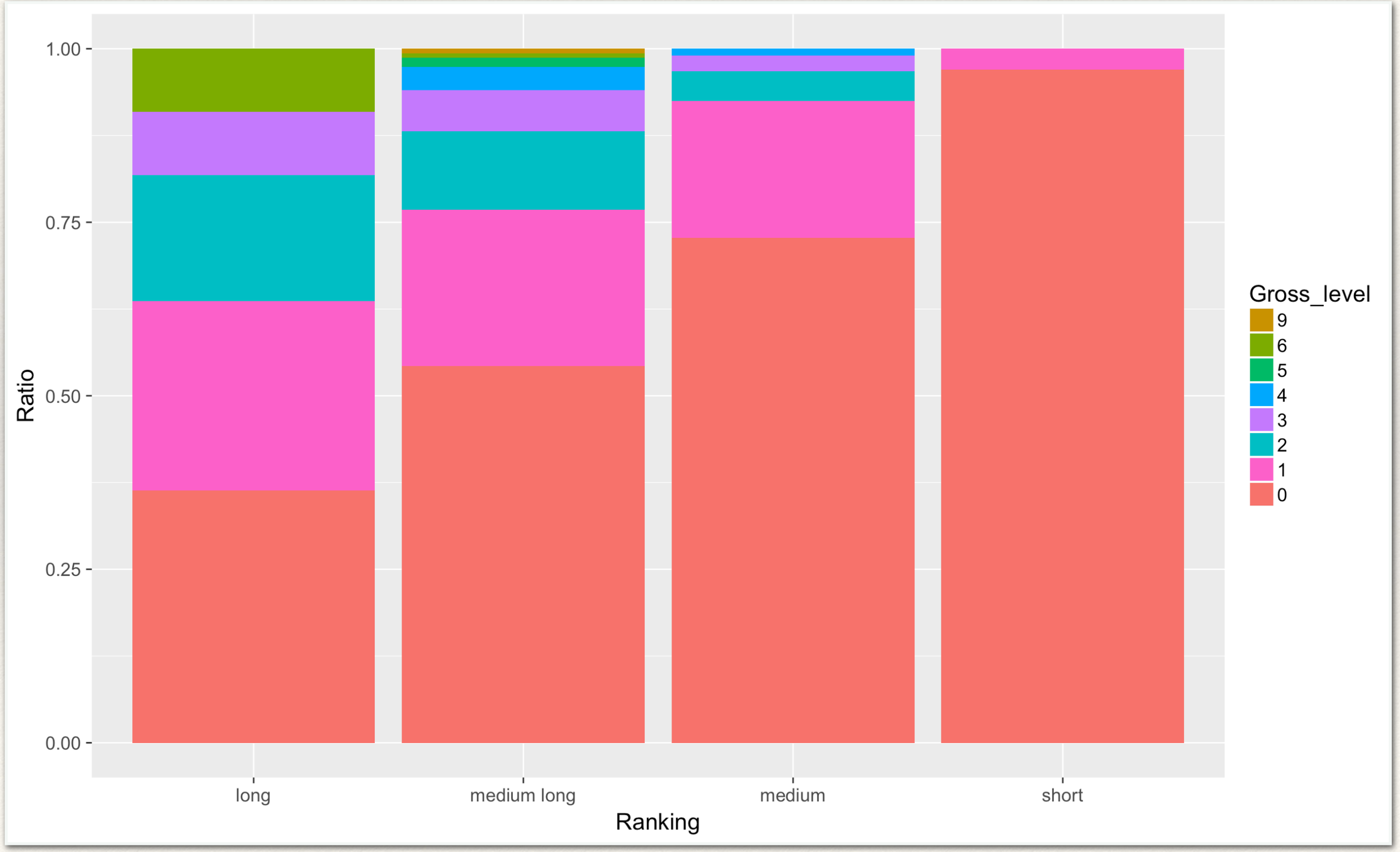

Running time Distribution
50
40
30
20
10
0
Count
90
120
150
180
Running Time


Running_Length Distribution
0/500
1
100
200
300
400
Running.length
long
medium long
medium
short

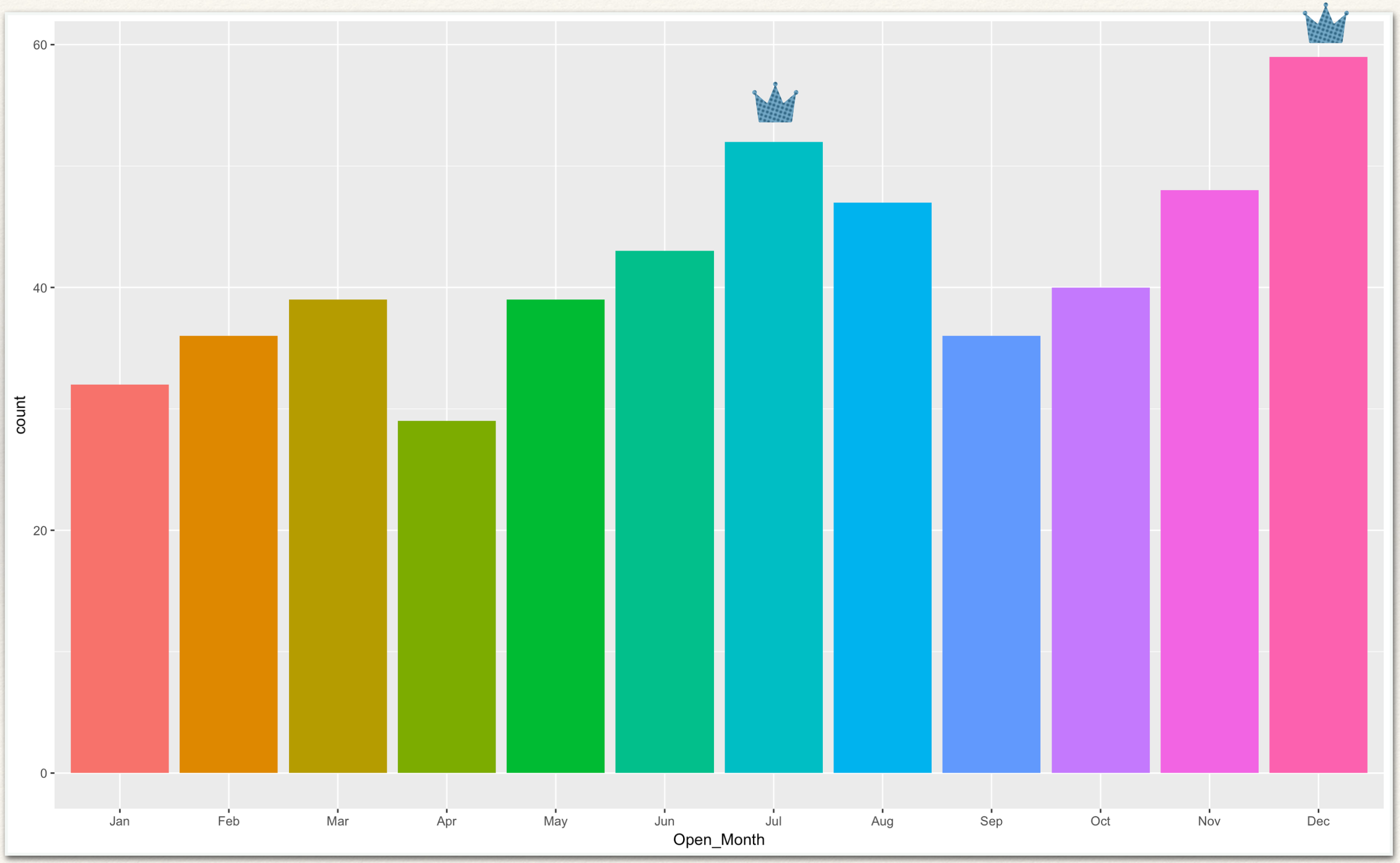
各電影片長排名分佈


1.00
0.75
0.50
0.25
0.00
Ratio
long
medium long
medium
short
Ranking
Ranking
1
2
3
4
5

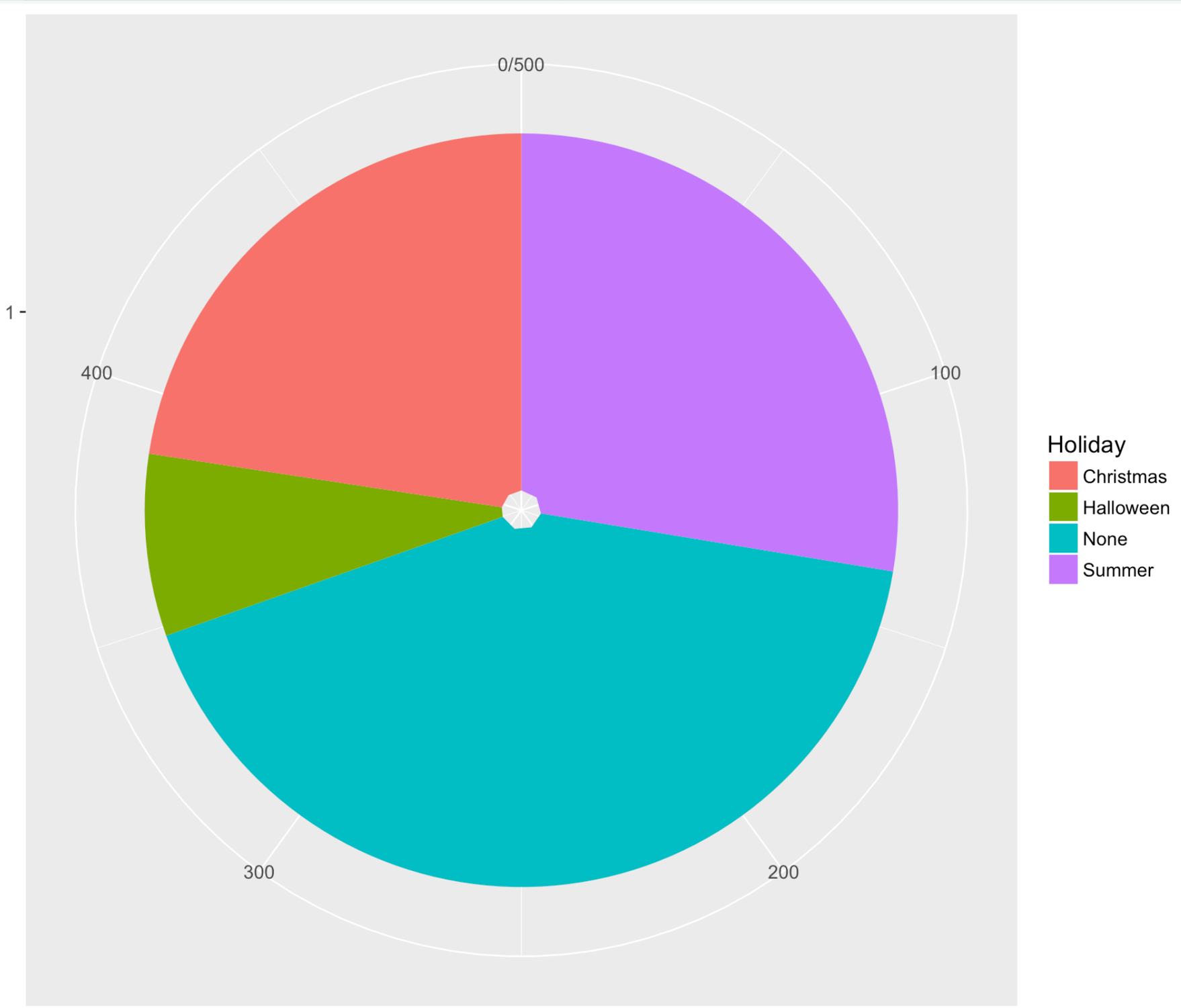
各電影片長總票房分佈


1.00
0.75
0.50
0.25
0.00
Ratio
long
medium long
medium
short
Ranking
Gross_level
9
6
5
4
3
2
1
0

各月份上映電影數


60
40
20
0
count
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Open_Month

不同節日電影上映比例



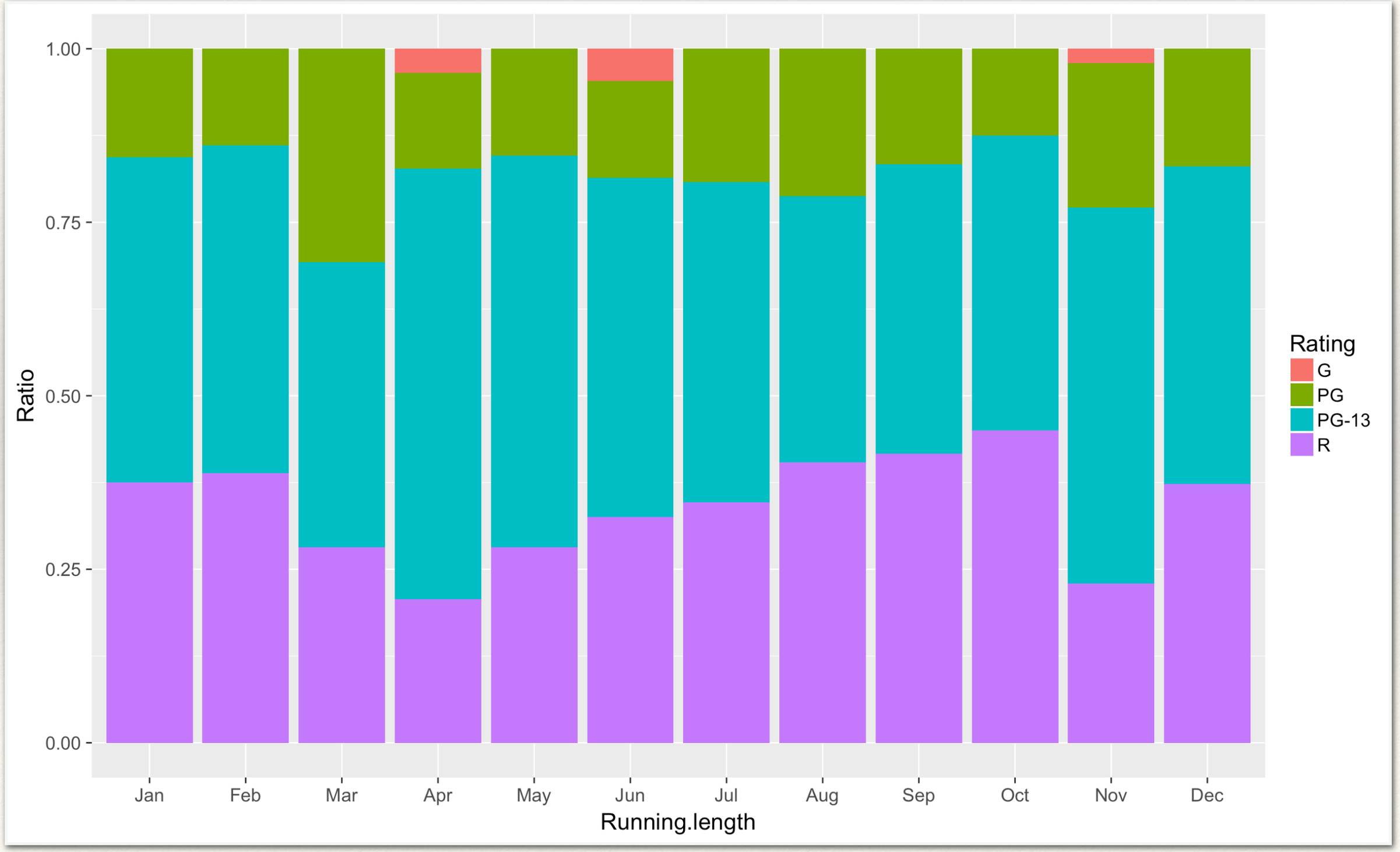
	Holiday	Month_Percentage
	<chr>	<dbl>
1	Christmas	0.226
2	Halloween	0.0780
3	None	0.420
4	Summer	0.276

不同節日電影上映比例

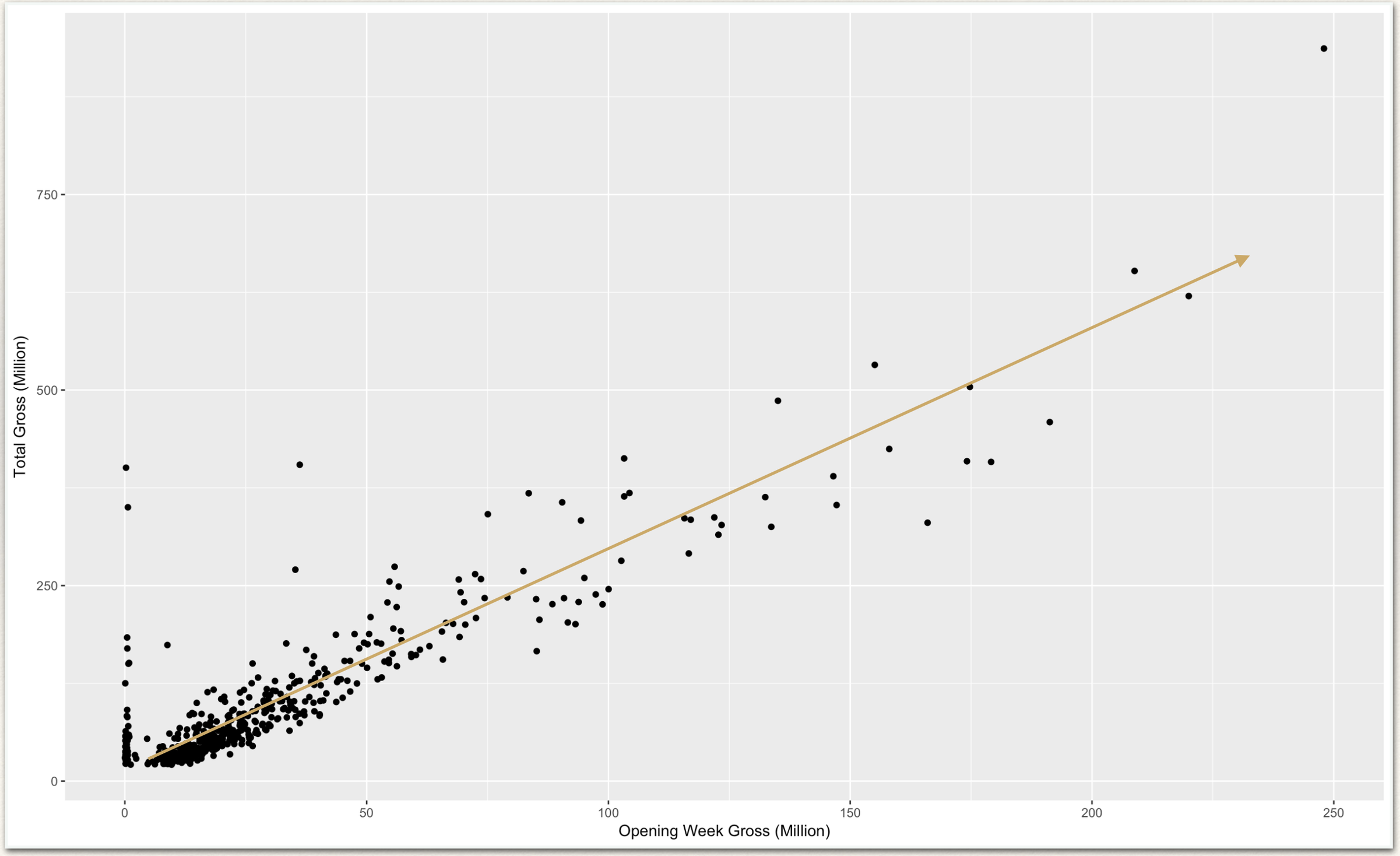
	Holiday	Month_Percentage
	<chr>	<dbl>
1	Christmas	0.226
2	Halloween	0.0780
3	None	0.420
4	Summer	0.276

	Type	Number	Summer_Percentage	Christmas_Percentage	Halloween_Percentage
	<fct>	<dbl>	<dbl>	<dbl>	<dbl>
1	Action	121.	0.347	0.124	0.0413
2	Adventure	61.	0.311	0.197	0.
3	Animation	53.	0.358	0.226	0.0377
4	Comedy	120.	0.292	0.242	0.0500
5	Crime	9.	0.111	0.333	0.222
6	Drama	104.	0.212	0.385	0.0865
7	Family	11.	0.273	0.273	0.0909
8	Fantasy	18.	0.278	0.333	0.111
9	Foreign	1.	1.00	0.	0.
10	Horror	46.	0.283	0.0217	0.217
11	Music	6.	0.333	0.667	0.
12	Romance	6.	0.167	0.	0.167
13	Sci-Fi	41.	0.268	0.171	0.0732
14	Thriller	43.	0.256	0.0465	0.140
15	War	3.	0.333	0.333	0.333

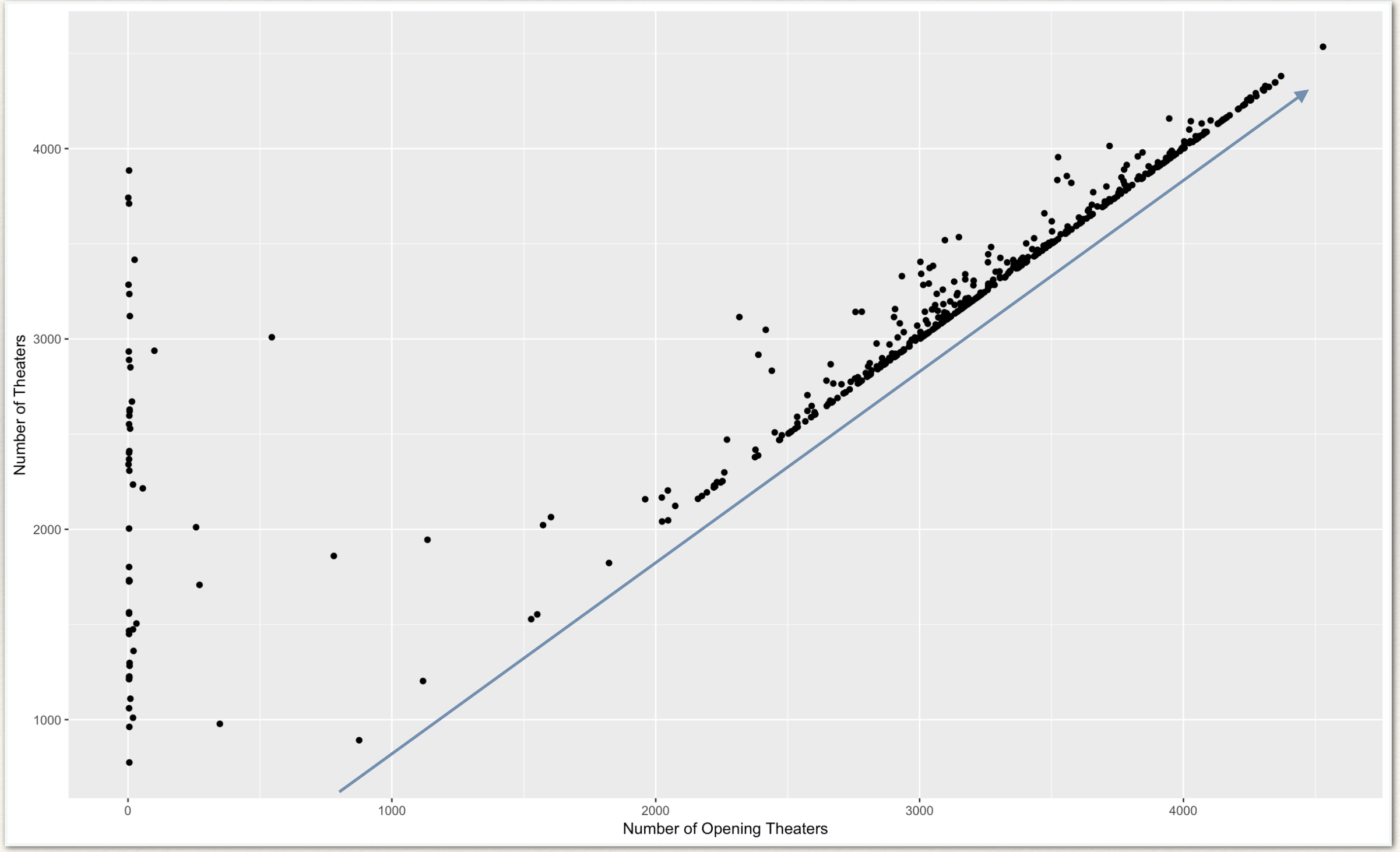
各月份上映電影分級分佈



上映週票房和總票房的關係



上映週和總上映影城數的關係



多變數線性迴歸模型

- 用2013到2016的資料，當作線性迴歸模型的Train，2017的資料當作被預測的Test

```
dataTrain = subset(data,Year != 2017)
dataTest2017 = subset(data,Year == 2017)
```

- 分析目標：用一部電影的各種資料，來預測其總票房

多變數線性迴歸模型

- 分析流程：
  - 一、選定哪些是適合的變數
  - 二、預測結果
  - 三、觀察預測的準確性

多變數線性迴歸模型

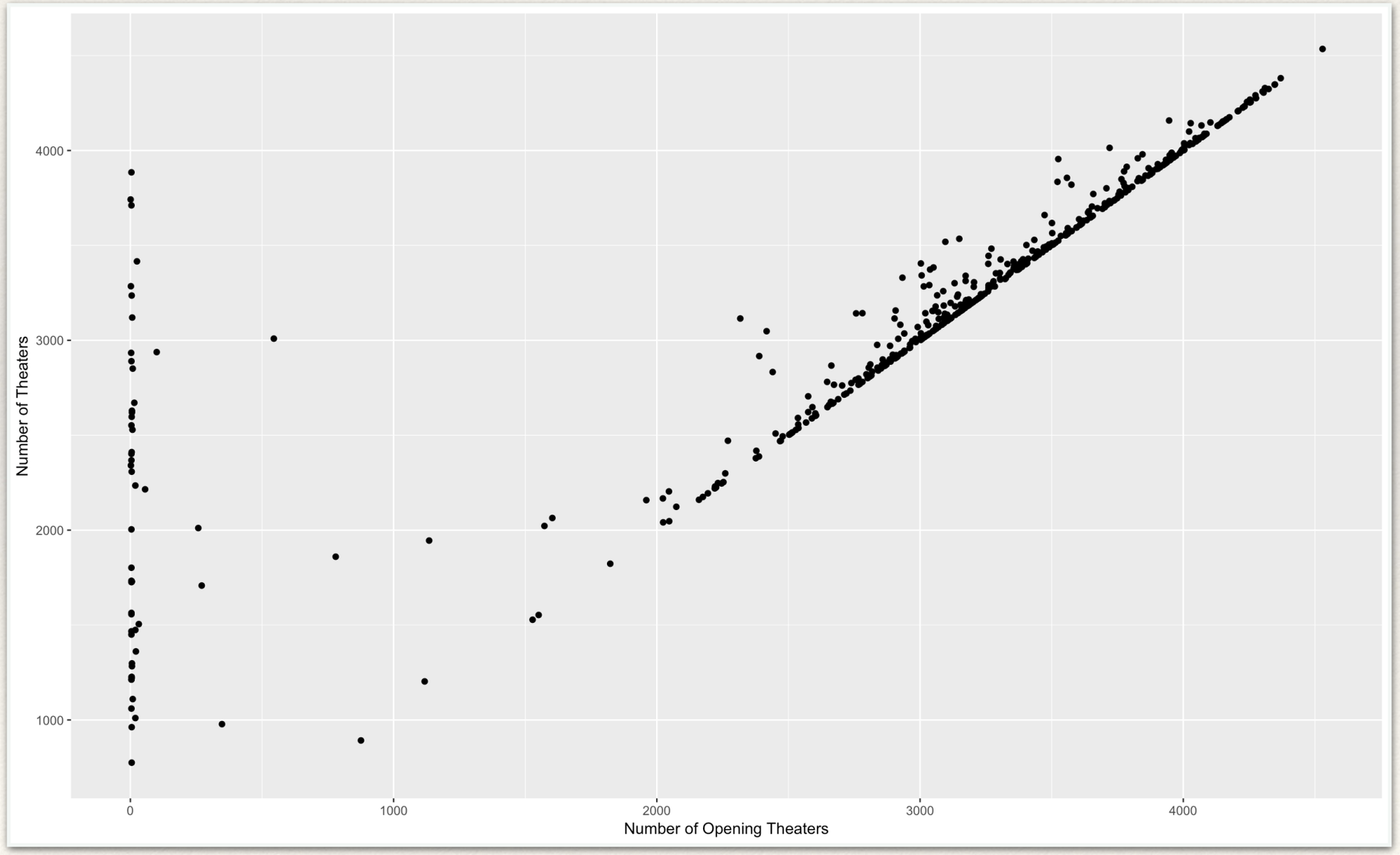
- 應變量：總票房
- 可使用的自變量：開幕票房、總上映戲院數、開幕上映戲院數、上映年份、上映月份、總上映天數、電影長度、電影分級、電影種類、片商

多變數線性迴歸模型：變數選擇

- 預測時，若使用總上映天數、總上映戲院數這兩個變數，似乎不太合理
- 此兩個變數的確切值，無法在電影下片前得知，所以這樣的模型只有在下片後才能跑，利用價值不大
- 此兩個變數的數值高，可能是總票房好的結果，而非原因

多變數線性迴歸模型：變數選擇

- 回顧：上映戲院數有時候會有大幅成長的現象



多變數線性迴歸模型：變數選擇

- 製造一個虛構（dummy）變數：
  - 「是否上映戲院數有大幅成長」
  - 把總上映戲院數 / 開幕上映戲院數 > 25 ，作為一個鑑別上映戲院數快速成長的虛構變數
  - 虛構變數：若符合此條件則為1，否則為0

多變數線性迴歸模型：變數選擇

- 先嘗試性地對所有變數做線性迴歸
  - 應變量：總票房
  - 自變量：開幕票房、開幕上映戲院數、上映年份、上映月份、電影長度、電影分級、電影種類、片商、是否上映戲院數有大幅成長

多變數線性迴歸模型：變數選擇

- 接著挑出統計上顯著（p-value=0.05為門檻）的變數，為以下幾項：
  - 開幕票房
  - 開幕上映戲院數
  - 是否上映戲院數有大幅成長
  - 開幕月份（為12月的時候）
  - 片商（為BV、Fox、LGS、Sony、STX、Tris、WB的時候）
  - 電影種類（為Animation、Sci-Fi Fantasy、Sci-Fi Thiller的時後）

多變數線性迴歸模型：變數選擇

- 為以上幾個統計上顯著的屬質變量，分別製造虛構變數
- 再用這些統計上顯著的變量，重新做一次線性迴歸

多變數線性迴歸模型：變數選擇

```
lm(formula = Total_Gross ~ Opening_Gross + Number_of_Opening_Theaters +
    Large_theater_growth + Open_Month_Dec + Studio_BV + Studio_Fox +
    Studio_LGS + Studio_Sony + Studio_STX + Studio_TriS + Studio_WB +
    Genre_Animation + Genre_SciFiFantasy + Genre_SciFiThriller,
    data = data2Train)

Residuals:
       Min         1Q     Median         3Q        Max
-131178772  -14752179   -1836604   11680023  271252241

Coefficients:
                            Estimate Std. Error t value Pr(>|t|)
(Intercept)                1.496e+07  1.095e+07   1.367 0.172500
Opening_Gross              2.788e+00  7.022e-02  39.699  < 2e-16 ***
Number_of_Opening_Theaters -6.920e+03  3.896e+03  -1.776 0.076461 .
Large_theater_growth        4.017e+07  1.259e+07   3.190 0.001541 **
Open_Month_Dec              3.856e+07  5.804e+06   6.644 1.05e-10 ***
Studio_BV                   2.806e+07  6.504e+06   4.315 2.03e-05 ***
Studio_Fox                  7.885e+06  5.930e+06   1.330 0.184383
Studio_LGS                  9.968e+06  9.023e+06   1.105 0.269958
Studio_Sony                 1.105e+07  6.652e+06   1.660 0.097644 .
Studio_STX                  2.551e+07  2.009e+07   1.270 0.204804
Studio_TriS                 1.085e+07  1.132e+07   0.958 0.338504
Studio_WB                   1.317e+07  5.212e+06   2.527 0.011912 *
Genre_Animation             4.562e+07  6.356e+06   7.178 3.66e-12 ***
Genre_SciFiFantasy          9.398e+07  2.531e+07   3.714 0.000234 ***
Genre_SciFiThriller         2.576e+07  1.743e+07   1.478 0.140173
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 34430000 on 385 degrees of freedom
Multiple R-squared:  0.8889,    Adjusted R-squared:  0.8849
F-statistic: 220.1 on 14 and 385 DF,  p-value: < 2.2e-16
```

多變數線性迴歸模型：變數選擇

- 反覆嘗試去除一些已經不再統計上顯著的變數（可能影響力已經被其他變數所解釋）
- 觀察到開幕上映戲院數的負面效果

多變數線性迴歸模型：變數選擇

- 最後選定的變數：
  - 應變量：總票房
  - 自變量：開幕票房、是否上映戲院數有大幅成長、是否上映月份為12月、是否片商為BV、是否電影種類為Animaiton、是否電影種類為Sci-Fi Fantasy

多變數線性迴歸模型：變數選擇

```
Call:
lm(formula = Total_Gross ~ Opening_Gross + Large_theater_growth +
    Open_Month_Dec + Studio_BV + Genre_Animation + Genre_SciFiFantasy,
    data = data2Train)

Residuals:
       Min          1Q      Median          3Q         Max
-122024956  -15384882   -2279130   10042454  278194740

Coefficients:
                      Estimate Std. Error t value Pr(>|t|)
(Intercept)          1.474e+06  2.653e+06   0.556 0.578729
Opening_Gross        2.716e+00  5.795e-02  46.871  < 2e-16 ***
Large_theater_growth 5.623e+07  6.582e+06   8.543 2.92e-16 ***
Open_Month_Dec       3.925e+07  5.730e+06   6.850 2.86e-11 ***
Studio_BV            2.208e+07  6.173e+06   3.577 0.000391 ***
Genre_Animation      4.210e+07  5.982e+06   7.038 8.76e-12 ***
Genre_SciFiFantasy   9.459e+07  2.534e+07   3.732 0.000218 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 34640000 on 393 degrees of freedom
Multiple R-squared:  0.8852,    Adjusted R-squared:  0.8835
F-statistic: 505.3 on 6 and 393 DF,  p-value: < 2.2e-16
```

多變數線性迴歸模型：預測結果

- 拿線性迴歸模型所估計的參數，跑2017年的電影資料

```
linearMod2.overall.refined2 <- lm(data=data2Train, Total_Gross ~ Opening_Gross + Large_theater_growth +
    Open_Month_Dec + Studio_BV + Genre_Animation + Genre_SciFiFantasy)
```

```
data2Test2017$Predicted_Total_Gross<-predict(linearMod2.overall.refined2, data2Test2017)
```

多變數線性迴歸模型：預測結果

```
> print(data2Test2017[c("Title", "Predicted_Yearly_Ranking","Predicted_Total_Gross")])
                                                Title Predicted_Yearly_Ranking Predicted_Total_Gross
401                                Star Wars: The Last Jedi                     1              754991233
402                             Beauty and the Beast (2017)                     2              498214119
405                          Guardians of the Galaxy Vol. 2                     3              421506972
408                                          Thor: Ragnarok                     4              356955941
407                                                      It                     5              336663956
406                                     Spider-Man: Homecoming                  6              319345632
403                                             Wonder Woman                    7              281927044
412                                The Fate of the Furious                      8              269799804
410                                          Justice League                     9              256369598
411                                                   Logan                     10             241619703
409                                         Despicable Me 3                     11             240318178
421                                                  Cars 3                     12             211482312
413                                                   Coco                      13             203643128
419 Pirates of the Caribbean: Dead Men Tell No Tales                          14             194630546
416                                The LEGO Batman Movie                      15             187540714
417                                         The Boss Baby                      16             179922925
420                                   Kong: Skull Island                       17             167232384
422                         War for the Planet of the Apes                     18             154296320
404                         Jumanji: Welcome to the Jungle                     19             138968258
414                                               Dunkirk                      20             138679633
427                                  Fifty Shades Darker                       21             128069473
425                          Transformers: The Last Knight                     22             122834856
438                                       Ferdinand                            23             119224074
437                                Power Rangers (2017)                       24             110938443
436                                     The Emoji Movie                       25             110206033
423                                               Split                       26             110152608
443              Captain Underpants: The First Epic Movie                     27             108357968
433                         Kingsman: The Golden Circle                       28             107469088
500                                 The Disaster Artist                       29             100244614
442                                     Alien: Covenant                       30              99694240
447                               The LEGO Ninjago Movie                      31              99072693
```

多變數線性迴歸模型：準確性觀察

- 比對預測的總票房和實際上的總票房，兩者的相關係數為0.93

- 預測目標一：預測票房達到 150 million 門檻的電影清單
- 預測目標二：預測年度前 20 名的電影清單

一、預測票房達到 150 million 門檻的電影清單

- 準確度：預測會達到 150 million 的18部電影中，總共有17部的實際票房，真的達到了 150 million 的門檻
- （沒達到門檻的有1部，其實際票房為 146 million ）

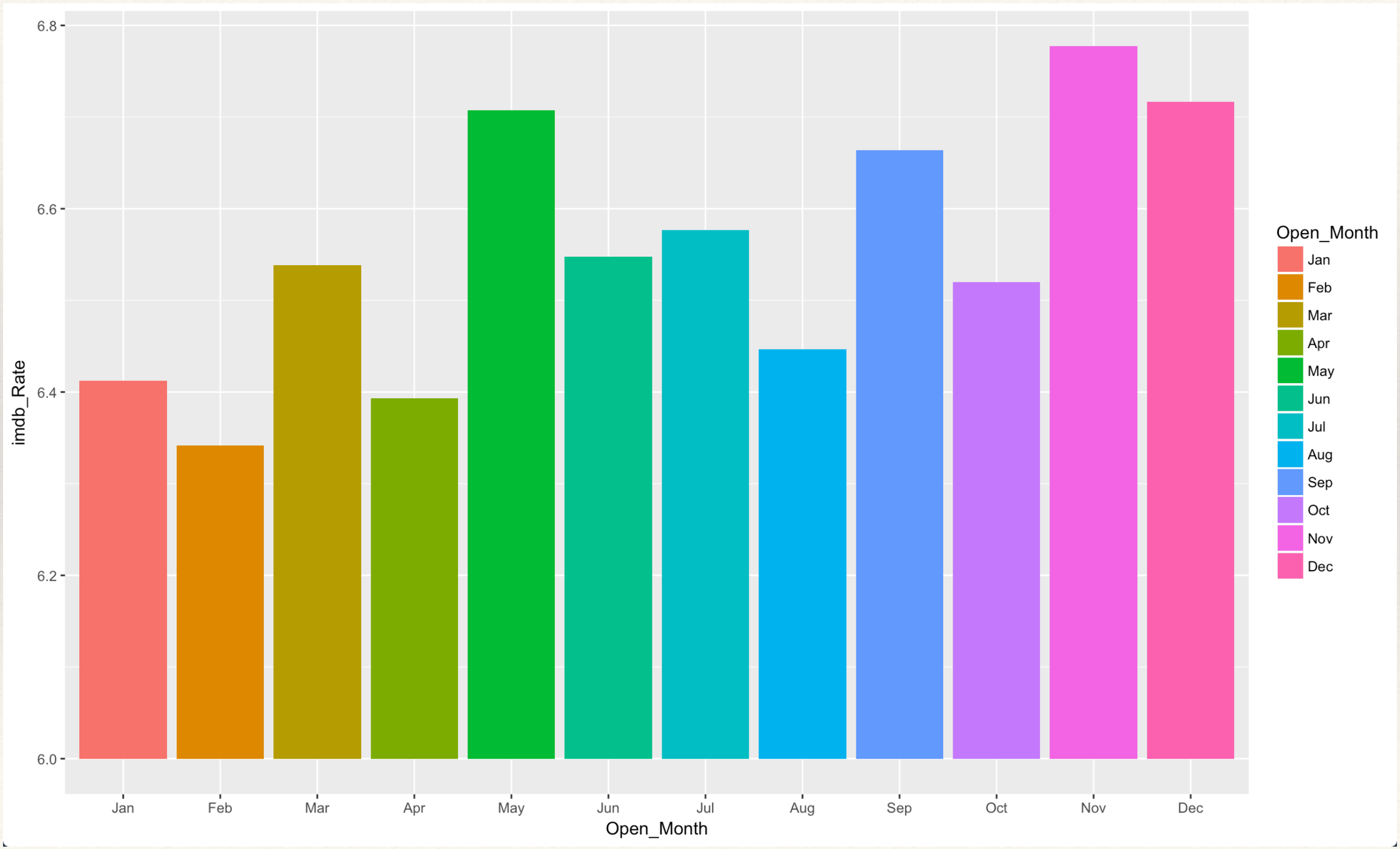
	Title	Predicted_Total_Gross	Total_Gross
401	Star Wars: The Last Jedi	754991233	620181382
402	Beauty and the Beast (2017)	498214119	504014165
405	Guardians of the Galaxy Vol. 2	421506972	389813101
408	Thor: Ragnarok	356955941	315058289
407	It	336663956	327481748
406	Spider-Man: Homecoming	319345632	334201140
403	Wonder Woman	281927044	412563408
412	The Fate of the Furious	269799804	226008385
410	Justice League	256369598	229024295
411	Logan	241619703	226277068
409	Despicable Me 3	240318178	264624300
421	Cars 3	211482312	152901115
413	Coco	203643128	209726015
419	Pirates of the Caribbean: Dead Men Tell No Tales	194630546	172558876
416	The LEGO Batman Movie	187540714	175750384
417	The Boss Baby	179922925	175003033
420	Kong: Skull Island	167232384	168052812
422	War for the Planet of the Apes	154296320	146880162

二、預測年度前20名的電影清單

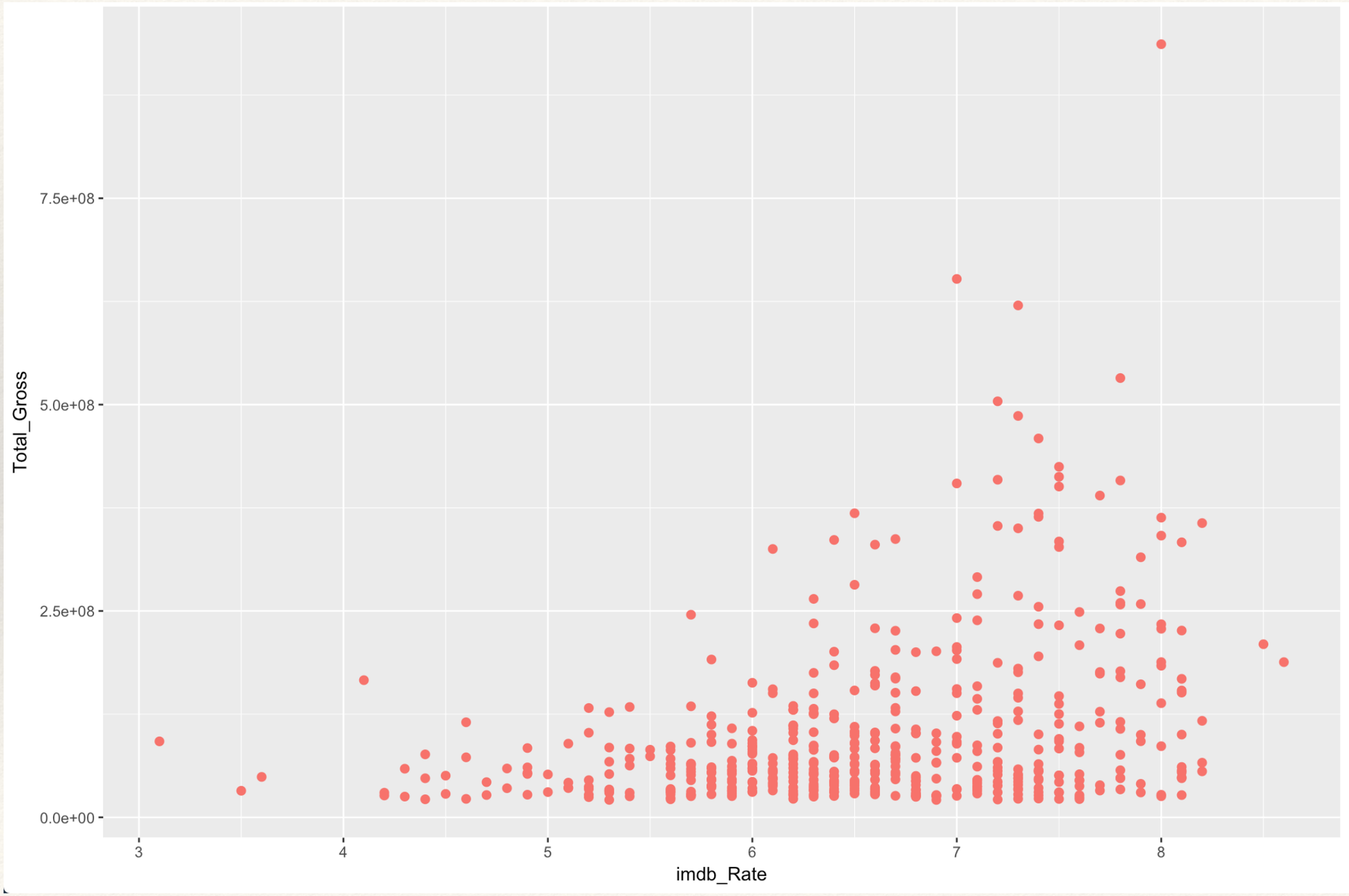
- 準確度：預測的年度前20名電影中，有18部真的進入了該年實際排名的前20名
- （沒進入前20名的有2部，實際排名分別為第21名、第22名）

	Title	Predicted_Yearly_Ranking	Yearly_Ranking
401	Star Wars: The Last Jedi	1	1
402	Beauty and the Beast (2017)	2	2
405	Guardians of the Galaxy Vol. 2	3	5
408	Thor: Ragnarok	4	8
407	It	5	7
406	Spider-Man: Homecoming	6	6
403	Wonder Woman	7	3
412	The Fate of the Furious	8	12
410	Justice League	9	10
411	Logan	10	11
409	Despicable Me 3	11	9
421	Cars 3	12	21
413	Coco	13	13
419	Pirates of the Caribbean: Dead Men Tell No Tales	14	19
416	The LEGO Batman Movie	15	16
417	The Boss Baby	16	17
420	Kong: Skull Island	17	20
422	War for the Planet of the Apes	18	22
404	Jumanji: Welcome to the Jungle	19	4
414	Dunkirk	20	14

IMDb參數加入


6.8
6.6
6.4
6.2
6.0
imdb_Rate
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Open_Month
Open_Month
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec

各電影的IMDb分數與其票房



使用 xgboost 預測

- 用2013到2016的資料，當作xgboost模型的Train，2017的資料當作被預測的Test

```
# xgboost
data2 = subset(data,Year != 2017)
data3 = subset(data,Year == 2017)
```

xgboost 預測用變數

- 選擇做為預測的變數：製片公司、開幕票房、首映戲院數、片長、類型、分級、上映月份、IMDb分數
- 因為xgboost的變數皆要是數值型，所以我們在預測前有作轉換

xgboost 預測目標

- 預測目標一：預測票房達到 150M 門檻的電影清單
- 預測目標二：預測年度前 20 名的電影清單

預測票房達到150M的結果

- 準確度：預測會達到 150M 的21部電影中，全部的實際票房，都達到了 150M 的門檻

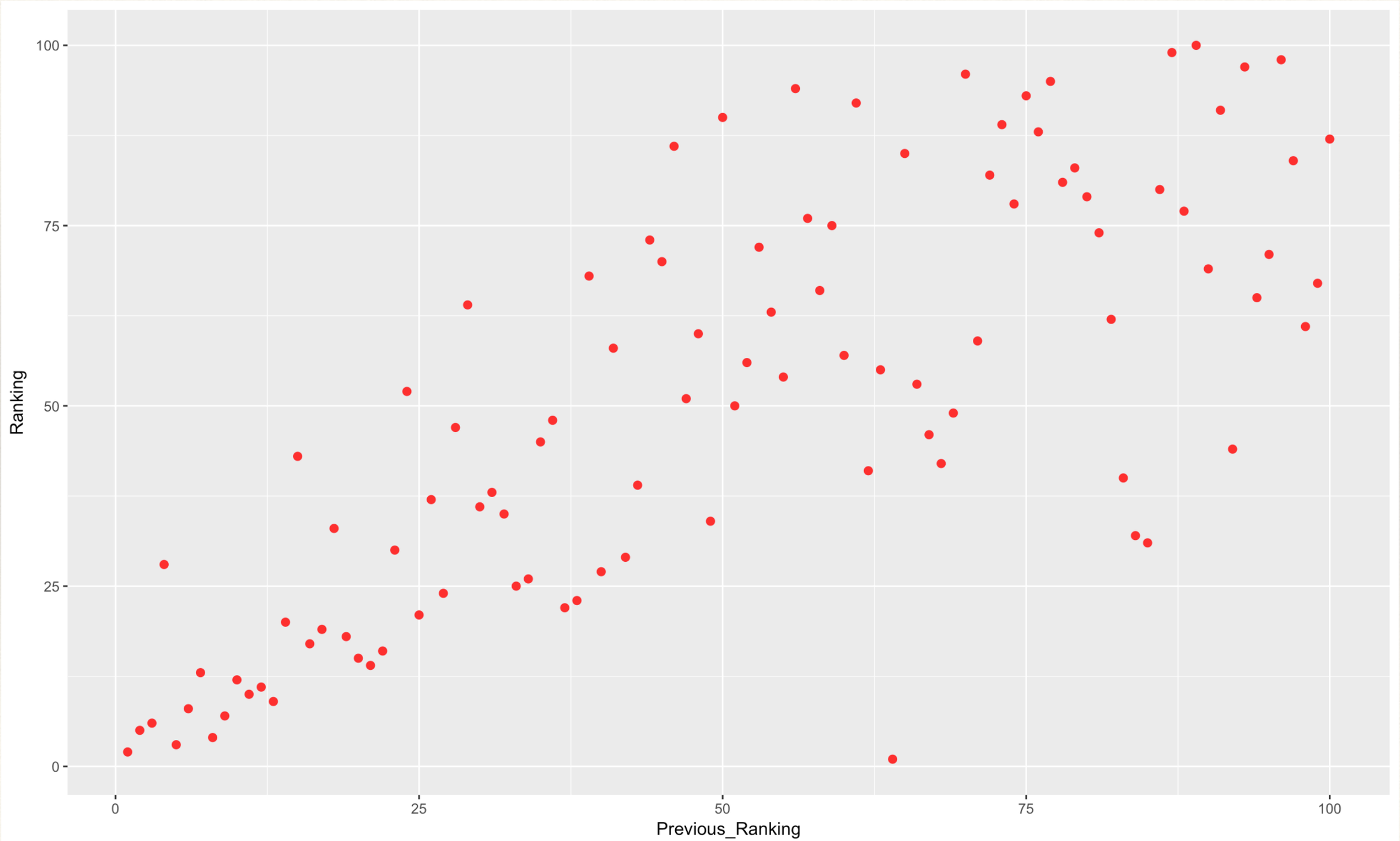
```
> subset(data6, Previous_Gross >= 150000000)
                                                Title Previous_Gross Total_Gross
1                                 Star Wars: The Last Jedi      620181382   634124288
2                            Beauty and the Beast (2017)      504014165   482559264
3                                          Wonder Woman      412563408   383158592
4                          Jumanji: Welcome to the Jungle      404515480   387792896
5                          Guardians of the Galaxy Vol. 2      389813101   406779264
6                                 Spider-Man: Homecoming      334201140   339689376
7                                                     It      327481748   336542752
8                                       Thor: Ragnarok      315058289   337063232
9                                     Despicable Me 3       264624300   263660912
10                                      Justice League      229024295   229074672
11                                                Logan     226277068   230771936
12                             The Fate of the Furious      226008385   228742896
13                                                Coco      209726015   199353920
14                                              Dunkirk     188045546   194746128
15                                              Get Out     176040665   175969056
16                           The LEGO Batman Movie          175750384   176151856
17                                    The Boss Baby         175003033   175917552
18                             The Greatest Showman         174016519   175017312
19 Pirates of the Caribbean: Dead Men Tell No Tales         172558876   167233120
20                                 Kong: Skull Island       168052812   169689456
21                                              Cars 3      152901115   152124880
```

預測年度前20名的結果

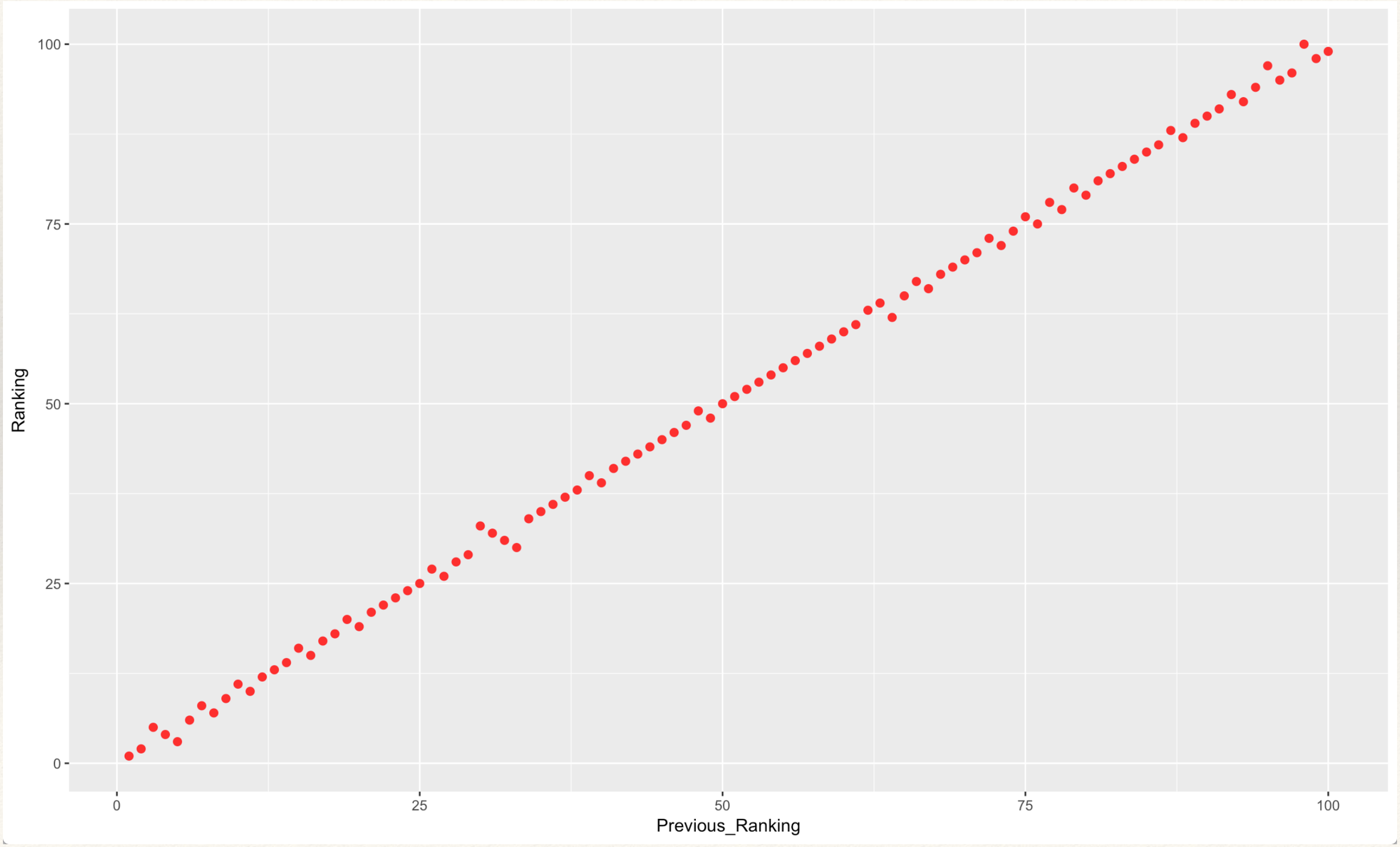
- 準確度：預測的年度前20名電影中，全部進入了該年實際排名的前20名
- 甚至有10部是直接命中！其他幾乎只有一部的差距

```
> subset(data6, Ranking <= 20)
   Ranking                                                Title
1        1                                Star Wars: The Last Jedi
2        2                             Beauty and the Beast (2017)
3        5                                         Wonder Woman
4        4                         Jumanji: Welcome to the Jungle
5        3                         Guardians of the Galaxy Vol. 2
6        6                                 Spider-Man: Homecoming
7        8                                                     It
8        7                                       Thor: Ragnarok
9        9                                     Despicable Me 3
10      11                                       Justice League
11      10                                                Logan
12      12                             The Fate of the Furious
13      13                                                Coco
14      14                                             Dunkirk
15      16                                            Get Out
16      15                              The LEGO Batman Movie
17      17                                      The Boss Baby
18      18                               The Greatest Showman
19      20 Pirates of the Caribbean: Dead Men Tell No Tales
20      19                                  Kong: Skull Island
```

預測年度前100名的結果



預測年度前100名的結果



xgboost 預測目標

- 雖然排名準確，但我們真正想知道的是預測票房的能力
- 我們拿預測的票房和實際票房相減，除以實際票房，得到誤差

```
> error <- mean(abs(data3$Total_Gross - data$Total_Gross[401:500] ) / data$Total_Gross[401:500])
> 100*(1 - error)
[1] 98.58225
```

- 準確率高達 98.58 % ！

檢討與改進

- 有更多變量可以使用，可能使得結果更好(ex:有名演員、導演的影響、imdb關注數量、評論的內容等等...)
- 爬蟲過程中，有少許資料遺失(人工補齊)

檢討與改進

- 進階目標：如果能在還沒上映就能準確預測首週和總票房
- 影響：可以讓電影公司決定如何制訂他們的行銷策略(如果上映前預測首週票房低於原先預期，可能需要更多廣告等等...)