

Stat 133 Final Project

Sol Ah Yoo & Sharon Hong

December 9, 2015

Objective - Analyze the correlation between GDP per year and the yearly box office total from 1980 to 2014 as well as top box office hits.

We want to see how the box office responds to the changes of the economy and how the popularity of the genres changed over the years. We also want to look at individual titles (the box office hit of the year) and analyze if there is a correlation between the amount it earned and the total box office of the year.

Libraries needed:

```
library(plyr)
library(ggplot2)
```

Data Sets Used:

- gdp_data - Data of GDP from 1980 to 2014 in the United States
- box_office - Data of yearly domestic box office totals
- top_movie - Data of yearly top grossing films in the US including genre, box office total, and production budget

```
top_movie <- read.csv("data/top_grossing_movie_yearly.csv")
```

```
box_office <- read.csv("data/box_office.csv")
```

```
gdp <- read.csv("data/gdp_data.csv")
```

```
head(top_movie)
```

```
##      Year                               Movie      Genre
## 1 1980 Star Wars Ep. V: The Empire Strikes Back Adventure
## 2 1981                               Raiders of the Lost Ark Adventure
## 3 1982                               ET: The Extra-Terrestrial Drama
## 4 1983      Star Wars Ep. VI: Return of the Jedi Adventure
## 5 1984                               Ghostbusters Comedy
## 6 1985                               Back to the Future Adventure
##      Production.Budget Total.Domestic.Box.Office Decade
## 1           23000000          290158751 1980s
## 2           20000000          248159971 1980s
## 3           10500000          434949459 1980s
## 4           32500000          309125409 1980s
```

```
## 5      30000000      242212467 1980s
## 6      19000000      212259762 1980s
```

`tail(top_movie)`

```
##      Year      Movie      Genre
## 30 2009      Avatar      Action
## 31 2010      Toy Story 3 Adventure
## 32 2011 Harry Potter and the Deathly Hallows: Part II Adventure
## 33 2012      The Avengers Adventure
## 34 2013      The Hunger Games: Catching Fire Adventure
## 35 2014      American Sniper      Drama
##      Production.Budget Total.Domestic.Box.Office Decade
## 30      4.25e+08      760507625 2000s
## 31      2.00e+08      415004880 2010s
## 32      1.25e+08      381011219 2010s
## 33      2.25e+08      623279547 2010s
## 34      1.30e+08      424668047 2010s
## 35      5.80e+07      350126372 2010s
```

`head(box_office)`

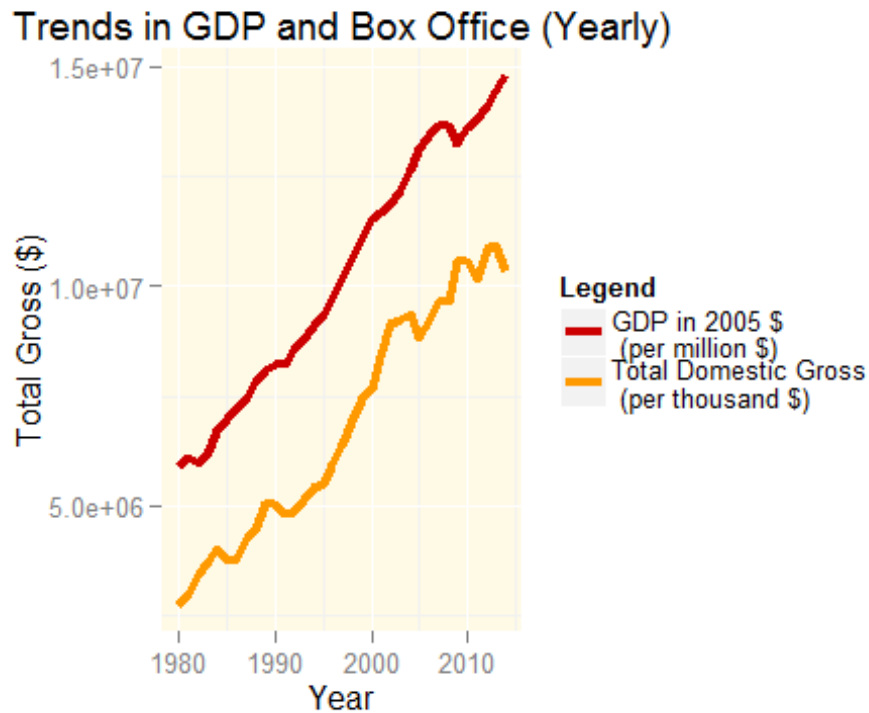
```
##      Year Total.Gross Change Tickets.Sold Change.1 X..of.Movies Total.Screens
## 1 1980 2749000000      NA 1022000000      NA      161      17590
## 2 1981 2966000000      7.9 1067000000      4.4      173      18040
## 3 1982 3453000000     16.4 1175000000     10.1      428      18020
## 4 1983 3766000000      9.1 1197000000      1.9      495      18884
## 5 1984 4031000000      7.0 1199000000      0.2      536      20200
## 6 1985 3749200000     -7.0 1056100000     -11.9      470      21147
##      Avg.Ticket.Price
## 1      2.69
## 2      2.78
## 3      2.94
## 4      3.15
## 5      3.36
## 6      3.55
```

`tail(box_office)`

```
##      Year Total.Gross Change Tickets.Sold Change.1 X..of.Movies
## 30 2009 10595500000     10.0 1412700000      5.3      521
## 31 2010 10565600000     -0.3 1339100000     -5.2      536
## 32 2011 10174300000     -3.7 1283000000     -4.2      602
## 33 2012 10837400000      6.5 1361500000      6.1      665
## 34 2013 10923600000      0.8 1343600000     -1.3      687
## 35 2014 10360600000     -5.2 1268100000     -5.6      698
##      Total.Screens Avg.Ticket.Price
## 30      NA      7.50
## 31      NA      7.89
## 32      NA      7.93
## 33      NA      7.96
```

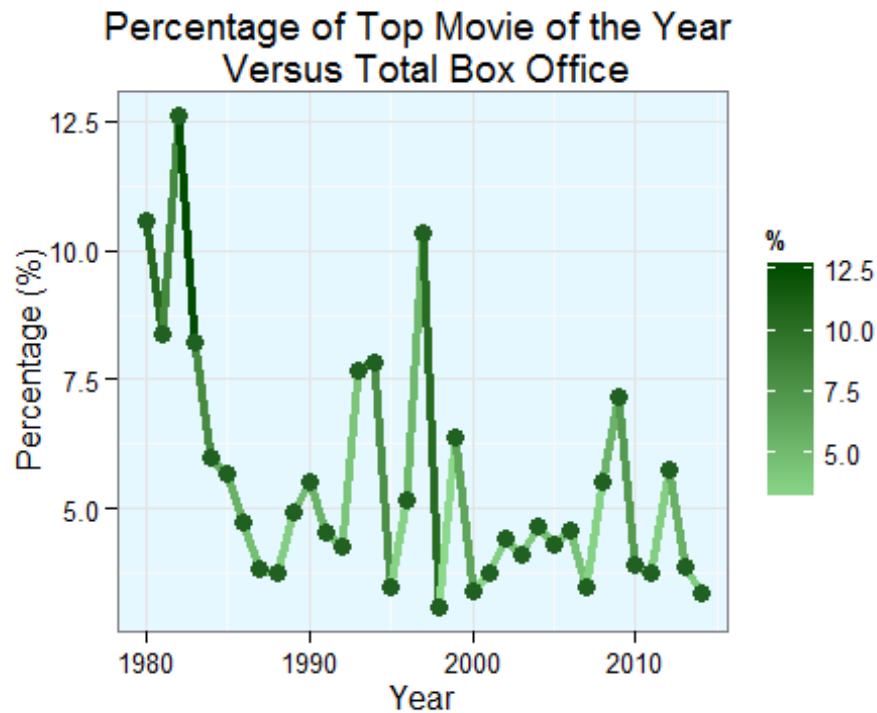
## 34	NA	8.13
## 35	NA	8.17

Is there a correlation between GDP and the gross domestic box office?



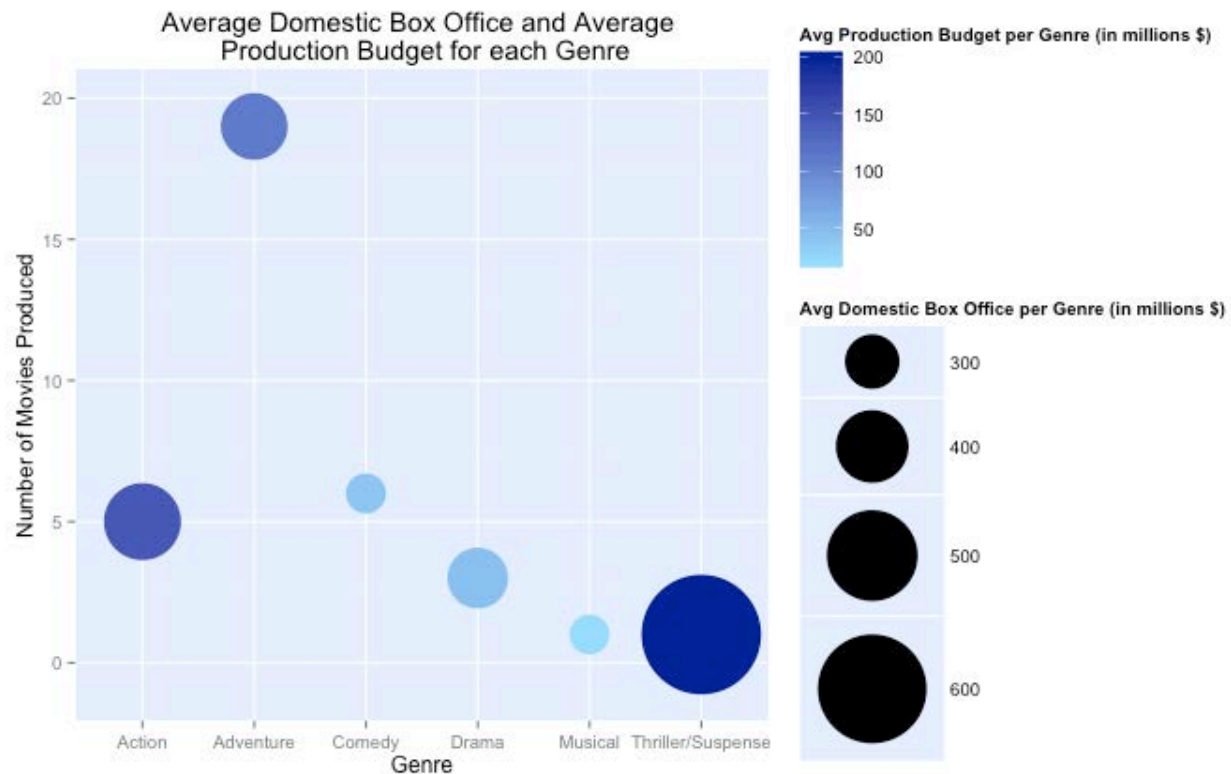
The GDP trend line shows fairly steady growth with a few notable exceptions, including the Great Recession of 2007-2009. The total domestic gross box office revenue generally has been following the same trends, but after the mid-2000's, we can see that there have been more fluctuations rather than a steady increase. The box office line is clearly decreasing, especially in comparison to GDP, which has been increasing, even though more movies are being produced yearly. This may be because of the rise of the age of the internet, which allows for people to stream movies (illegally) or use legal streaming services such as Netflix or Hulu, which allow for movies to be streamed with a set monthly subscription.

What is the relationship between the top grossing film of the year and the total box office?



The overall trend shows that a lower percentage of the total box office revenue comes from the top movie hit of the year. Only a few more recent hits, such as Titanic (1997) and Avatar (2009) break this trend with large spikes, with domestic box office totals of \$658,672,302 and \$760,507,625, respectively. After E.T. (1982), there is a huge decline in the percentage of the top movie as a share of the total box office and now more and more movies are produced yearly, which also decreases the share of each top hit as a percentage of the total box office. This shows that the higher spikes of Titanic and Avatar are more impressive because they outperformed a higher number of movies than earlier films.

Find the average total domestic box office by genre. We want to see how much each top movie (by genre) made in the box office compared to their production budget.



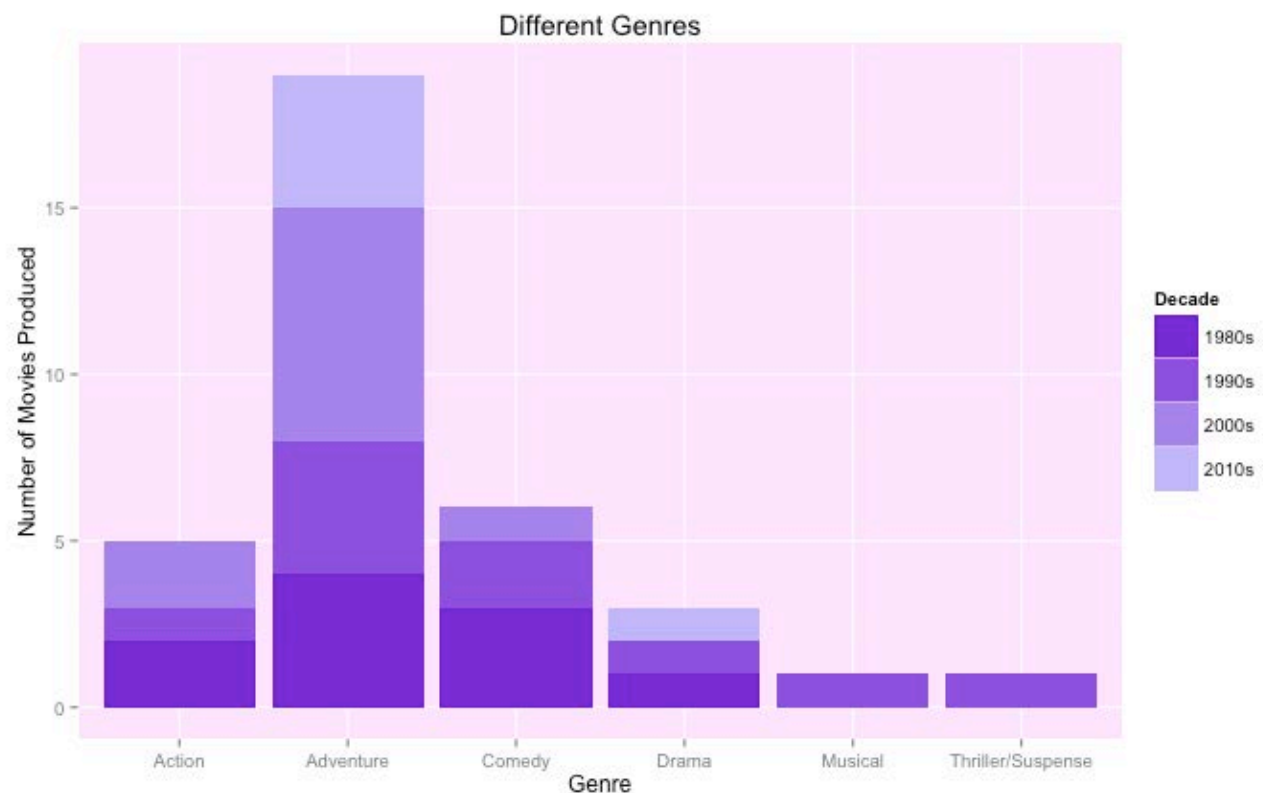
This is the analysis of the top grossing movies of the year with the x axis being the type of genre and the y axis being the frequency of each genre type by the top box office hits from 1980 to 2014.

The higher the bubbles are on the graph, the more movies of that genre were the top movies of the year. We can see that adventure movies are typically the most popular type of movies and that other types of genre, such as musicals and dramas are usually not the top movie of the year.

Also, the size of the bubbles represents the average box office total of each type of genre. For instance, because there was only one immensely popular thriller/suspense ("Titanic") in the top from 1980 to 2014, the area of that circle is disproportionately larger than the other circles, whereas because there were 19 top adventure movies in last 35 years, the bubble for the adventure genre accurately represents what a typical popular adventure film grosses.

Additionally, the colors represent the average production budget of the movies grouped by genre. The darker the color, the higher the budget and vice versa. You can see that action, adventure, and thriller movies tend to have a higher budget than the other genres because those genres tend to use more special effects than genres such as drama and comedy. They also tend to require more elaborate costumes and background props, which increase the budget. In addition, they tend to make more in the box office, given that the circles are larger for the action, adventure, and thriller movies than for the comedy and drama movies.

How many different genres are there and what decade did they come out in? What are the most popular genres?



Adventure movies tend to be more popular with the general public, while comedy comes in second, followed closely by action movies. The least favorites are musicals and thriller/suspense.

Adventure movies have been the most popular type of genre in the 2010s. With the advancement of technology, movie makers have been able to make more realistic looking scenes for movies that require intensive computer-generated imagery, such as superhero movies and fantasy movies. Now, filmmakers are able to pull off the effects required for the

movie to seem realistic to the audience. The last couple of top movies (The Hunger Games, Avengers, Harry Potter, Avatar, etc) have all been filled with CGI and other effects, which if they were made in the 80's, might not have been as popular. Also, adventure movies are liked by all types of people, old or young, boy or girl, while other genres like drama are more likely to be preferred by the older population than younger children.