Part 2:

In this section, we explored how genre affects the rate and box office of a film. Genre as one of the most important attributes of a film, has a strong link to the film’s box office and rate. Commercial films like Action or Adventure films, would earn money than other films nowadays. While the overall rate of this kind of profit-directed film is always a little lower than art films like Drama or Music film. However, the released year of a film would also affect the box office, hence we have taken year into account.

This part chooses the top 30 films’ data for recent 20 years from IMDB website, including each film’s name, release year, rate, and box office. The box office and rate of each year and each genre have also been calculated.

**Data：**

Data source: All needed data in this part is grabbed from the website and after processing, all data is integrated into a csv file which contains 3 attributes: name, rate and box office.

Data structure: Box office, rate, released year, and genre are all attributes of a film, and since when we searched data, we chosen top 30 films in box office and then compared with their rates, we transform data in the following data structure:

**Software:** D3 and Excel

Using Excel to process csv file to get the proper data and using D3.js to conduct visualization. Considering the data structure, we choose using tree view to visualize data in this part.

**Visualization Work:**

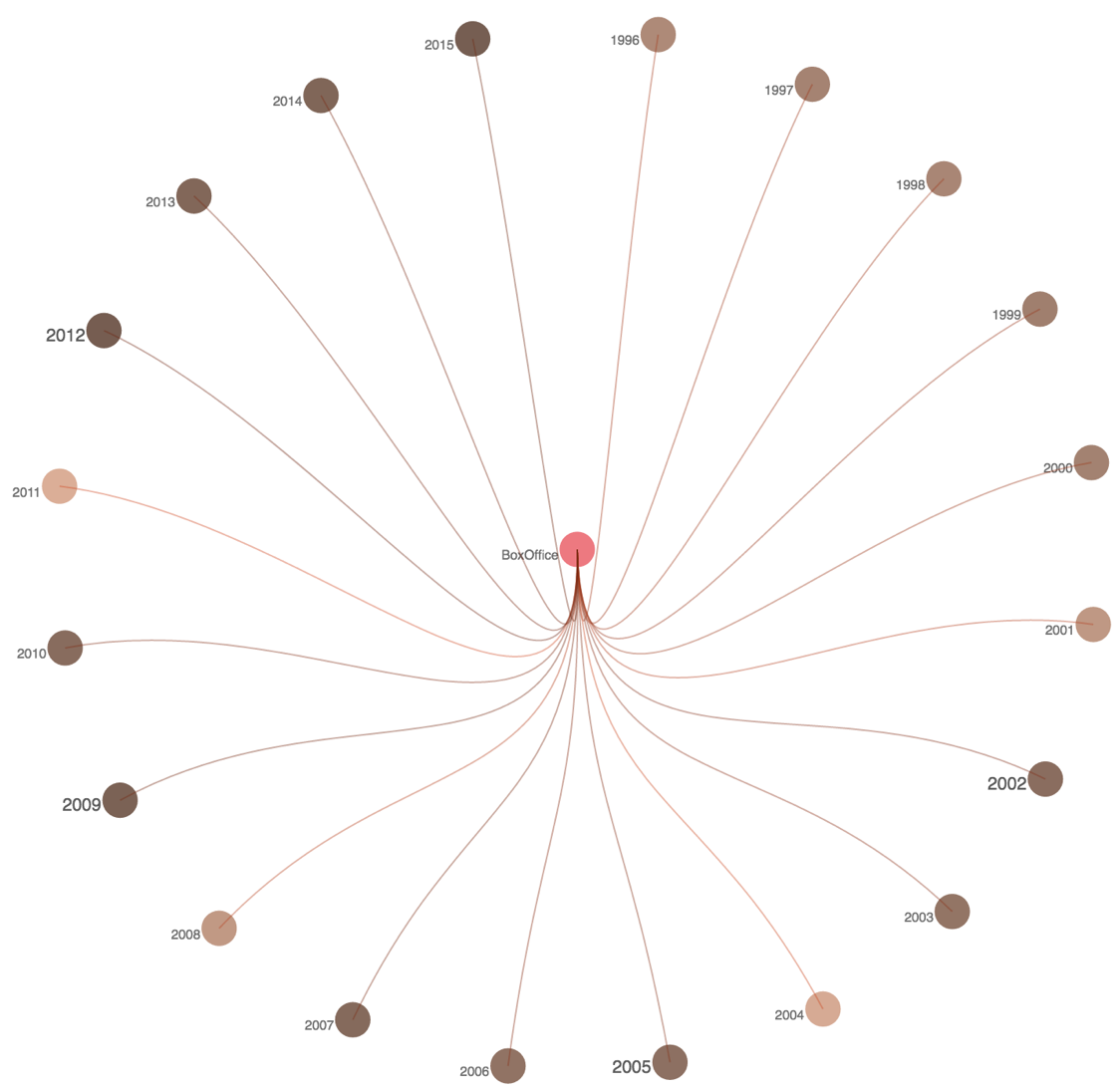


Image 1. Relationship between Box Office and Rate in different years

Image 1 is a tree view of different year’s box office, the data is from 1996 to 2015. The darkness of color represents the number of box office, the higher one year’s films’ total box office is, the darker the color will be. The size and darkness of label represent the the number of rate, if one year’s films’ average rate is higher, the label of it will be bigger and darker.

From this diagram, we can directly recognize that from 1996-2015, in spite of some special year such as 2004,2008 and 2011, the box office of each year increases. The color from 1996 to 2001 is relatively light, while from 2012 to 2015 the color has become quite dark. This indicates the development of film industry because of the improvement of people’s life.

Also, it is obviously that some years, for example year 2002 and 2011, did share a similarity at rate mark and box office. However, it is not always the case, since some years, such as 2015, get a high box office and at same time received a low rate.

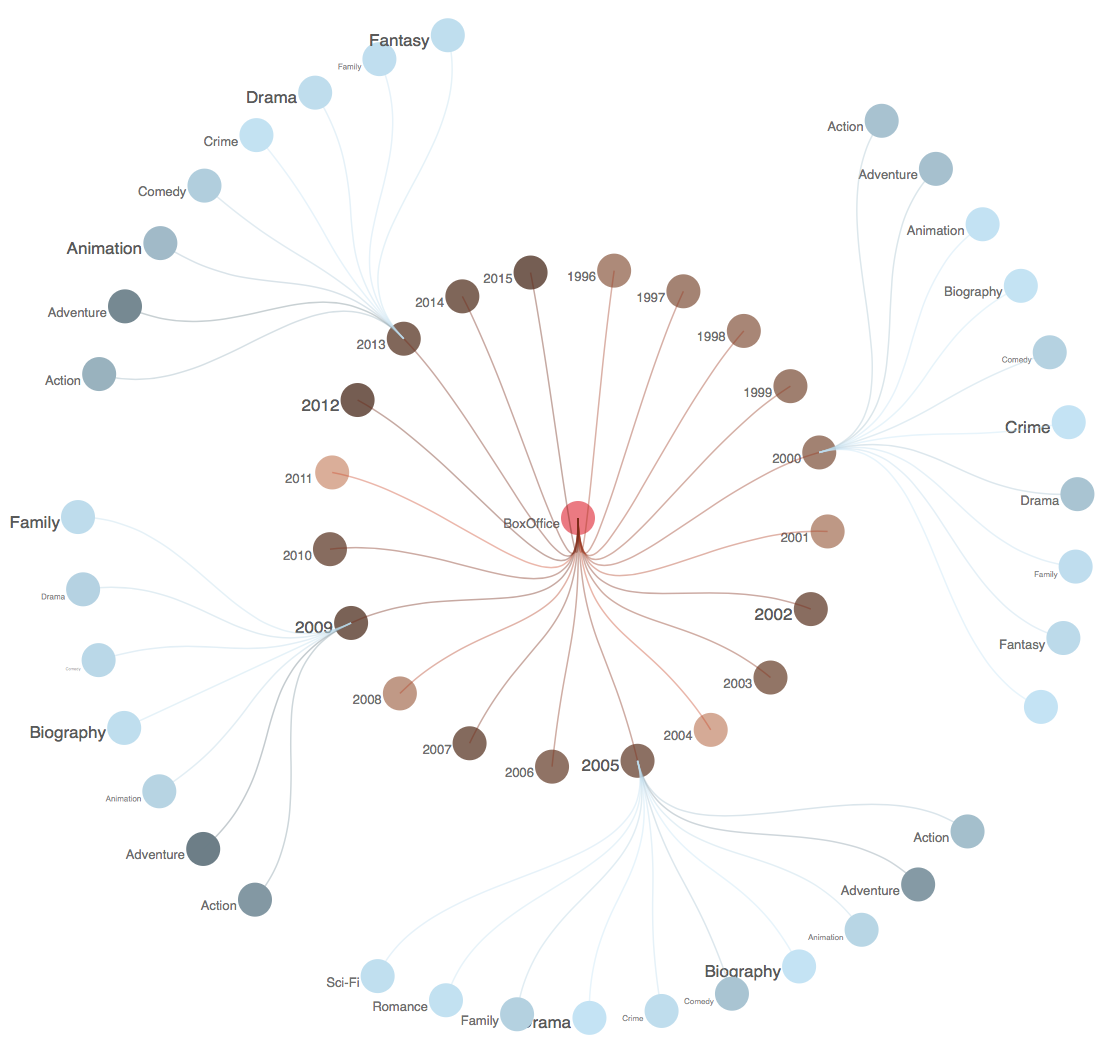


Image 2 Relationship between Box Office and Rate of different genres

After analyzing the relationship between box office and rate in different years, we can further explore that if this relationship would change when it comes to different genres as shown in Image 2.

After clicking different year points, the detail of this year’s box office and rate will be shown.

From this picture, by comparing the composition of box office and rate in year 2000,2005,2009 and 2013, we can figure out that genres like Adventure, Action, Animation and Comedy normally would earn a relatively high box office and a relatively low rate. On the contrary, genres like Crime, Documentary and Fantasy earn a good reputation rather than great money. Also, the rest genres like drama and Biography won both in box office and rate mark, Romance and Horror lost in both areas.

Another interesting point from this view is that if we pay attention to the top film genre in box office and rate each year, we could find that the top film genre in box office is almost the same as Action and Adventure. While the top genre in rate varies a lot, in 2005, the top genre is Biography, but in 2000, the top genre is Crime. Hence, using box office to evaluate a film may be easier and more data is needed when utilizing rate marks to evaluate a film.

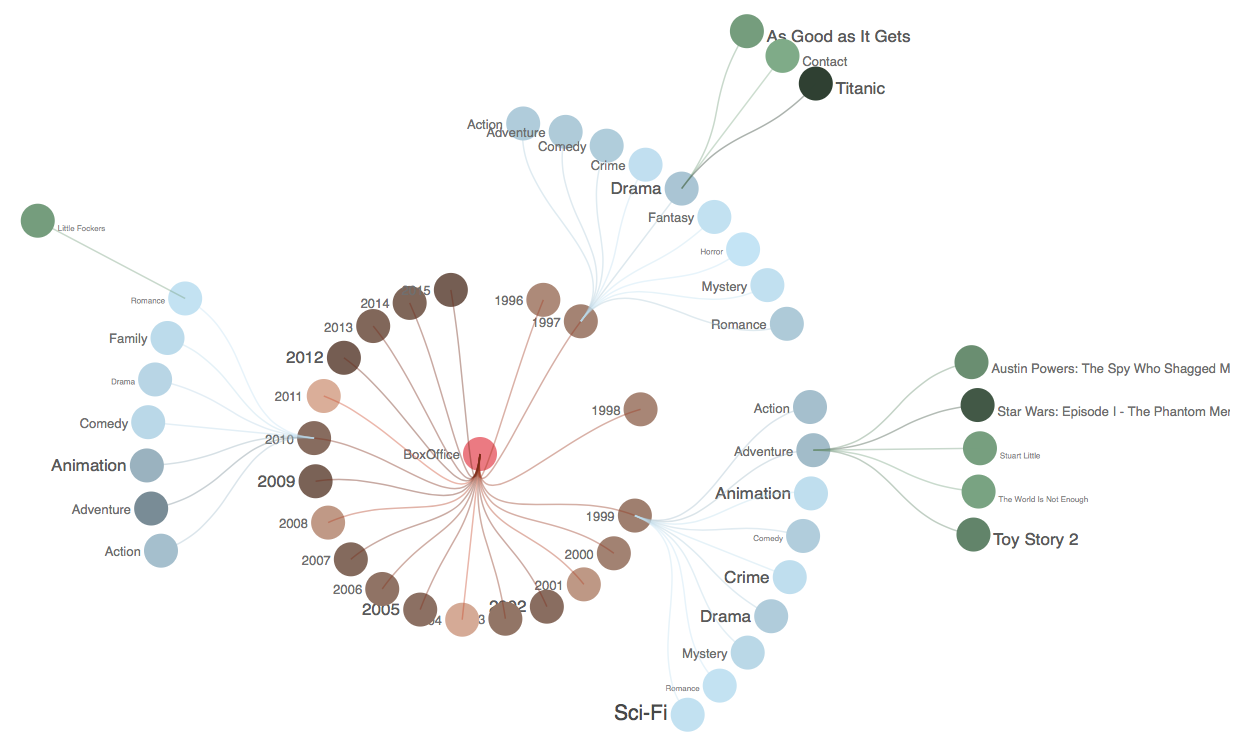


Image 3 Relationship between Box Office and Rate of different films

Genre is the second layer of this tree, then when click on each genre point, the view will illustrate the box office and rate of each film this genre contains. As shown in image 3, the fantasy genre in year 1997 owned the highest box office which is a little abnormal. And after clicking on it, we found the reason: the world famous film Titanic, it is this film that made fantasy genre become the top one in year 1997.However, the rate mark of Titanic is not quite high, which is a little confusing.

Adventure, one of the most popular genre, earned the highest box office in 1999.The 5 adventure films in 1999 all had outstanding performance, especially “Star Wars: Episode 1”. However, despite the fact that “Star Wars” sold quite more tickets than “Toy Story 2”, its rate mark is far lower than the other one which means “Toy Story 2” may has a more meaningful contents.

By comparing relationship between box office and rate of different films, we could further account for how genres affect film’s box office and reputation.

**Difficulties and Solving Methods**

To draw dynamic circle tree view need to solve two problems:

1.Traditional tree layout is up to down, therefore, the function provided in d3 would only return the x value and y value of each point in up to down layout.

2.Dynamic transformation of tree view contains hiding points and paths, increasing point and paths, dynamically moving points and paths to the right position, which is always changing, at the right time, and readjust the layout of tree to fit the increment or decrement of depth.

To solve the first problem, I use first figure out the returned x and y value meaning try to rewrite a new function using returned value as an input to get the new x and y values in circle layout.

By implementing enter, update and exit function and the reconstruction function of tree, I succeeded in solve the quite complicated second problem

**Limitation of D3**

D3 is really a powerful visualizing tool, however, since its big change from version 3 to version 4, plenty of functions change not only in the name but also the whole design logic. While version 4 is quite new that not so many tutorial materials can be found. And the official API document lacks details.