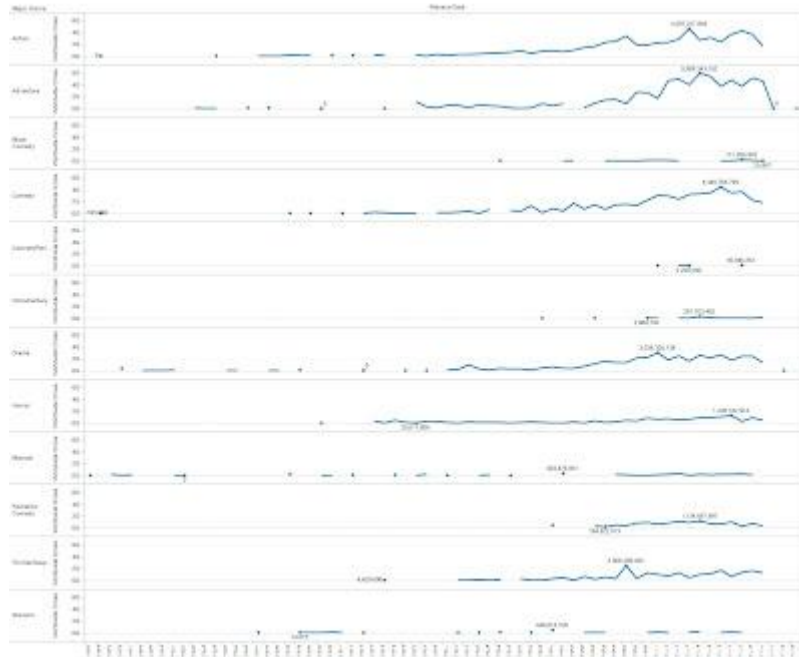


## Data Exploration

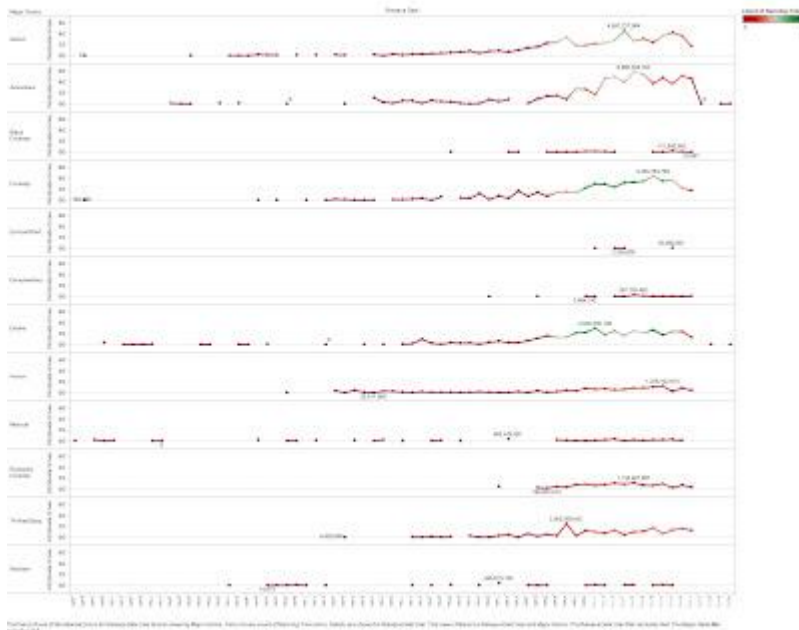
Initially based on my interest in movies, I chose movies excel file. Then by looking into it, I became eager to know answer of questions below:

Does the general taste of people about movies have changed through years and decades or not? If yes, were they attracted by violent movies in past time or funny ones? My guess is violent ones (because of difficult life situation in past time) but let's test it.



Worldwide gross per each genre through different years

This chart shows that general interest of people in movies has increased through years. Almost before 1980, people were not interested in movies, as we go forward they began to show their interest in many genres especially adventure and action, horror. If we go forward in years, we see that through years and decades, general interest in drama and horror remains constant while there is a big pick in adventure, action and comedy genre. In 2003-2004 years, people are mostly interested in Action and Adventure movies, however in 2006 their taste changed and they were more willing to comedy movies. Also it is observable that people generally are less attracted to Western, black comedy, concert/performance and documentary genre. Then I questioned myself Is general interest in a genre related to running time of its movies? If yes, how?

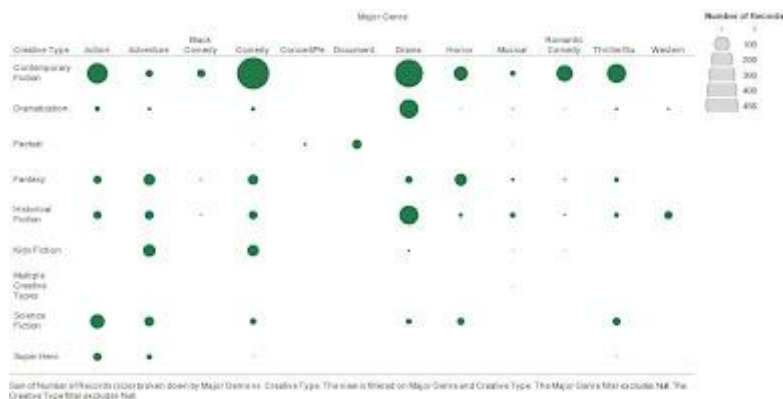


running time of different genre through years

The above diagram is the previous one which the parameter of running time of movies is added to. As running time of a movie is closer to 30 minutes, it is darker in green, and as it is closer to 0 minutes, it is darker in red.

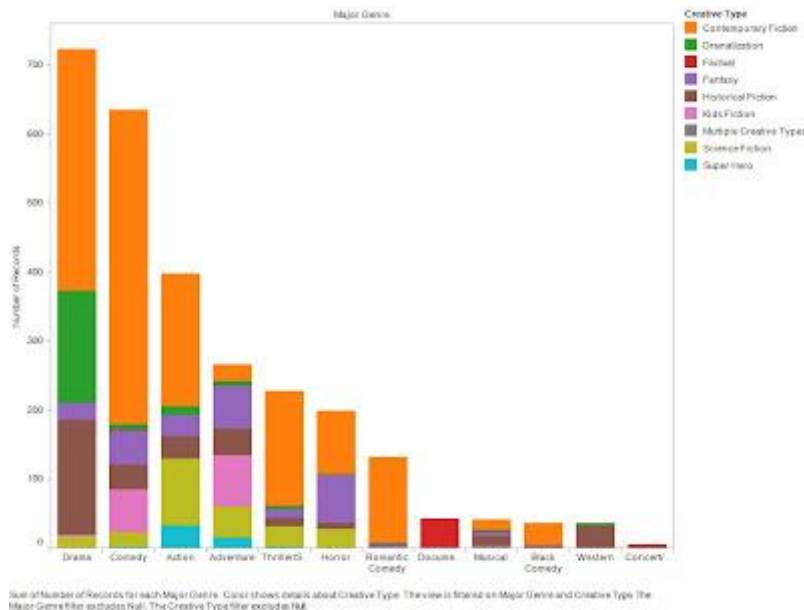
It is observable in past time, average running time of movies in all genre had been short but throughout the years it has been increased in some genres. Running time of comedy and drama movies has more increase as compared with other genres.

Also we can say that general interest in a genre is not related to running time of its movies, as we see interest in adventure genre in years after 2004 is rather high but their color is red (=their running time is below average), or interest in comedy genre in years between 1998 and 2002 is not much high but their running time is utmost. Then I decided to find out how many films are in each genre and each creative type, so I drew diagram below.



Number of movies in each genre-type

In above chart, size of a disk describes number of movies in a particular genre-type. It is shown that contemporary fiction has the largest distribution, specifically in comedy genre and after that drama genre. Multiple creative types has the lowest distribution and if we ignore it, factual and super hero have the lowest number of movies. For better comparisons we can take a look at bar chart version of this diagram:



Number of movies in each genre-type (bar chart)

This diagram is sorted along with major genre descending by number of records (which each record represent a movie). Here we can see clearly that drama genre constitute majority of movies, after that comedy has more number of movies. Also we can see that all documentary and concert genre movies are of factual type. So we can conclude drama and comedy genre have largest distribution of movies, and among creative type, contemporary fiction exists in most genre and consists the largest number of movies.

Then I was willing to check rating of genre against their distribution. I was curious if a genre/creative type exists that has a large number of movies but a low rating. So I added IMDB rating to the original disk diagram:



### Number of movies and their rating in different genre-type

In above chart, the red-grey-green bar indicates IMDB rate of specific a genre-type. It shows that IMDB rate of drama and comedy genres universally are better than other genres and some genres such as adventure or horror has a low IMDB rate. Then I understood that comparing rate of different genre in this way is not correct because rate of a genre is a function of number of movies in that genre (the more a genre has movies, the higher probability it has to have higher rate) so I decided to draw my diagram based on percentile not the exact number of movies.

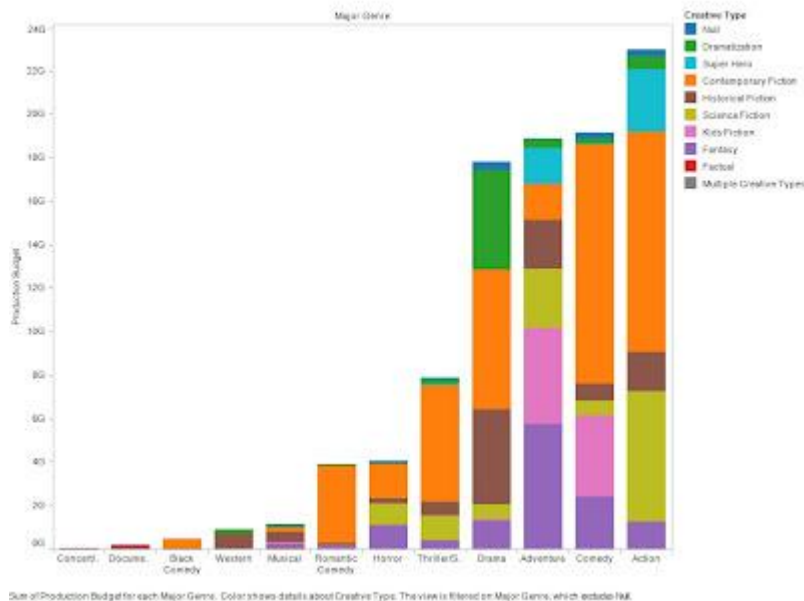
In diagram below, size of disk represents percentage of movies which are in a specific genre-type, and color of disk describe the difference between IMDB rate of that genre-type and average IMDB rate of all genre-types.



Here we can see that drama movies universally make a fair percentage of total (size of disk) but have low rates (dark red color). Also contemporary fiction movies make a large portion of movies (due to size of disks in first row) and do well in comedy, action, horror and romantic genres but do really badly in drama. Kids fiction does badly in adventure but not in comedy, average number of kids fiction movies in adventure and comedy genres are the same but they are more popular in comedy genre and it sounds reasonable because kids like comedy better than adventure. Historical

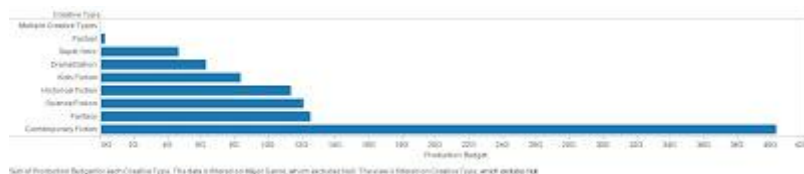
fiction and dramatization are the worst relative to others because they have large distribution (sum of disk sizes of their rows) but low rate (color of their disks are all red)

Then I became curious to know which genre-type most costly movies belong to. So I draw a diagram based on production budget of each genre-type.



Production budget of different genres

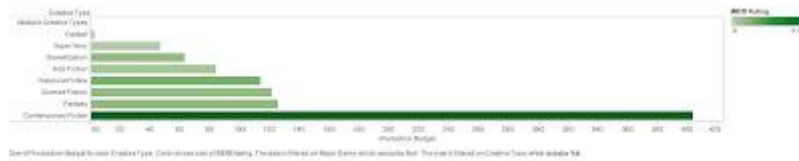
Diagram above shows production budget of different genre (each column represent a genre), columns are sorted increasingly based on their length. It is observable that concert/performance genre has the lowest budget and action genre has the largest budget. comedy, adventure and drama have the same and fairly a large budget. Contemporary fiction type exists in almost 75% of all genres and have the largest budget in 58% of them. Fantasy type also exist in 66% of all genres and is the second largest budget consumer. For an assessment on relation between budget and type, I draw the chart below:



Product budget of different creative types

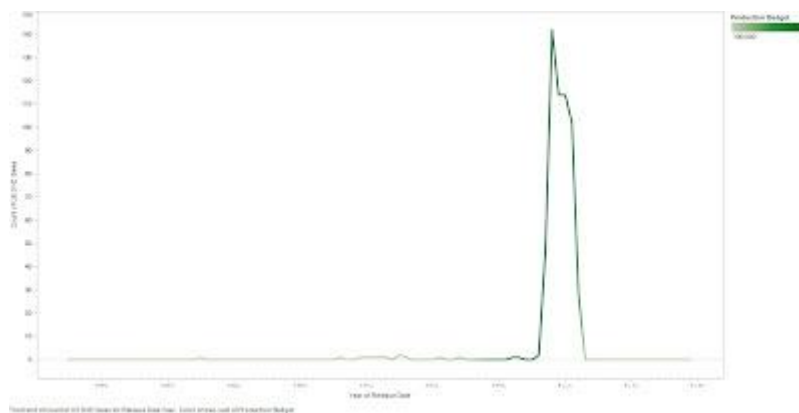
It shows fantasy, science fiction and historical fiction have a close budget amount, factual type almost has a zero budget. Then I wanted to check whether there is a relation between production budget and IMDB rate of a type or not. So I drew chart

beneath. It emphasizes that although fantasy, science fiction and historical fiction have nearly a same amount of budget, historical fiction has higher IMDB rate (darker green). Also we see that dramatization has lower budget than kids fiction but has higher rate. So there is not a relation between budget and rate of a type.



Budget and IMDB rate of different creative types

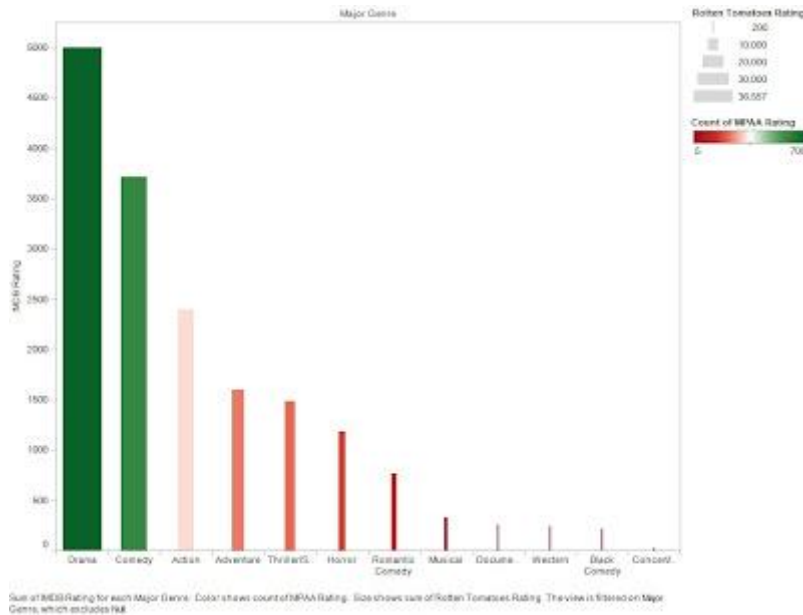
Then I questioned myself since which year, people were willing toward buying film dvds. Diagram below shows number of US dvd sale versus released date of films and their budget.



Budget and rate of dvd sale vs. released date

This diagram shows since 1998 and afterwards production budget of movies has increased significantly but their dvd sale rate was constant until 2004. In 2004, rate of dvd sales increased from 2 to 142 (a significant promotion). Although maximum dvd sale rate is at 2006 (after 2006 rate of sale has decreased), average production budget of movies was large till 2010. So we can say that between years of 1998 and 2004, film maker companies has had a profit issue because they had spent a large budget on their movies but obtained low dvd sale rate.

Then I thought that different rating system have various criteria, so maybe rating of IMDB, MPAA and RTR are different for some genres. So I drew diagram below:



Rating of IMDB, MPAA and RTR for different genres

The more a bar is tall, higher IMDB rate it has, the darker color of a bar is, higher MPAA rate it has, and the thicker a bar is, higher RTR rate it has. Diagram is sorted by IMDB rate. So we see that for drama genre all three rating system have had a high rate, while for action genre, IMDB and RTR have had a fairly high rate but MPAA has had a rate below average (That seems reasonable because MPAA rating is done by parents and if a movie consists of violent scenes, they will give it a low rate. Therefore action genre should have a low MPAA rate)

We see that MPAA and RTR have given nearly a same rank for adventure and thriller genres but IMDB has given a lower one to thriller. Also documentary, western, black comedy and concert have a same MPAA rate (their color are the same) and a same RTR rate (their thickness are the same) but they are not the same from IMDB opinion. IMDB rates movies based on users' votes and apparently users prefer documentary and western to concert performance.