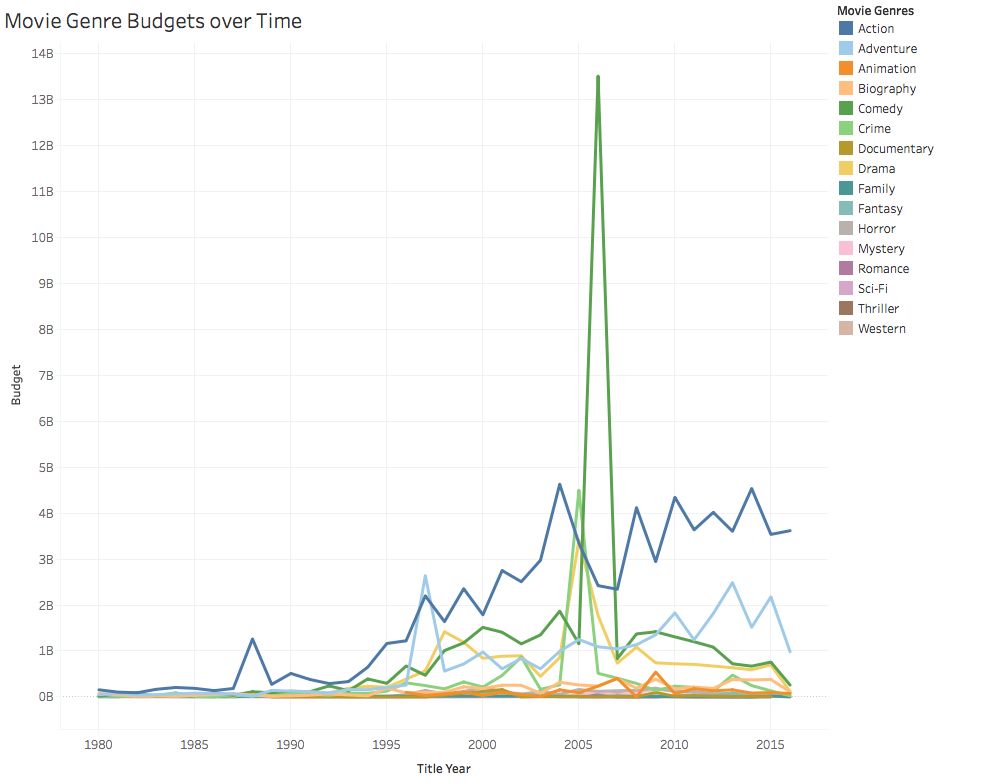
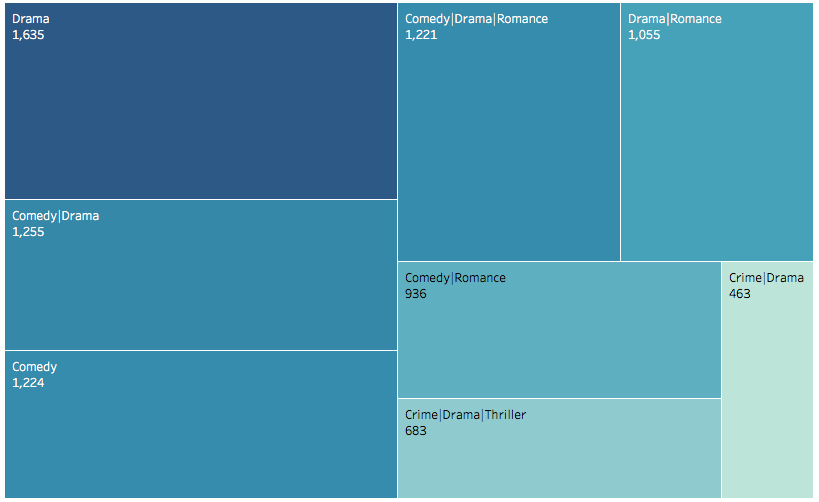
Some [background](https://www.kaggle.com/deepmatrix/imdb-5000-movie-dataset) on the IMDB 5000 dataset. Includes data on movies from 1916-2016.

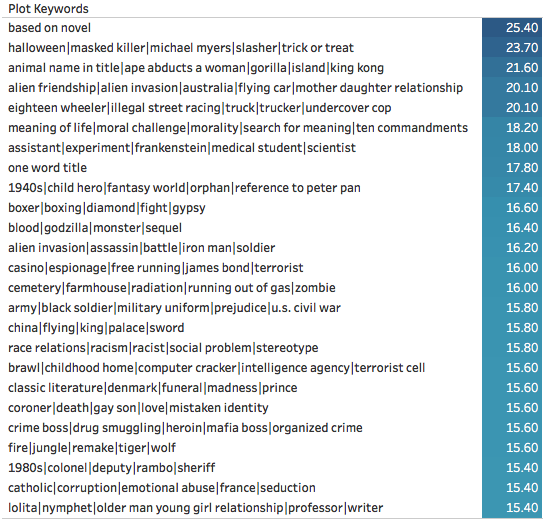
Movie genre budget have seen interesting trends over time-- noticeably, there’s a spike near 2007 for comedy movies. Meanwhile Dramas budgets saw a spike near 2005, but have been in decline since. Action movie budgets are on the rise...



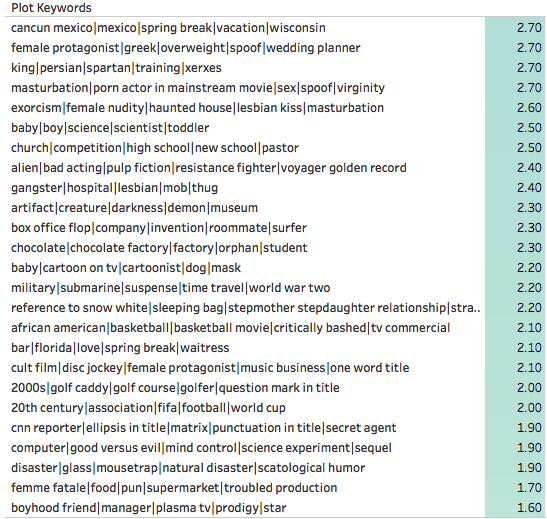
In contrast to the budget observation for the Drama genre, it is the highest rated genre amongst the top 8 genres as summed by IMDB Scores



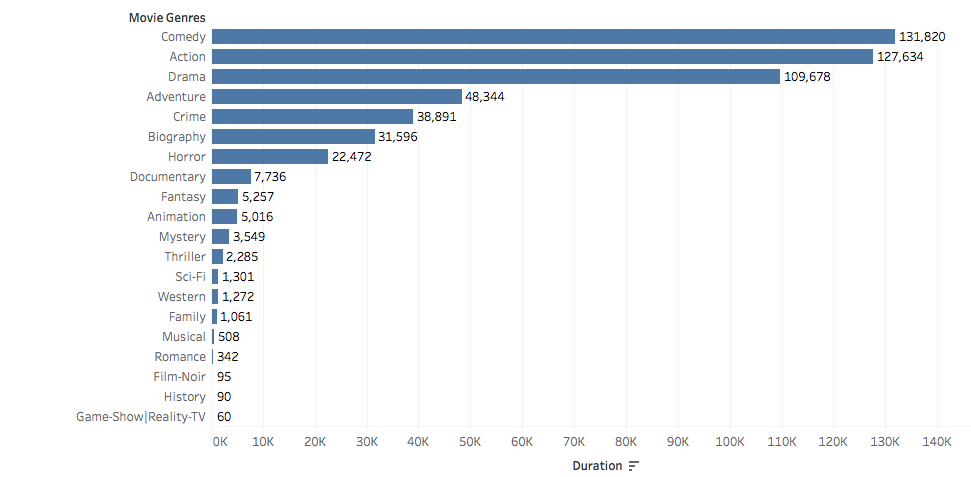
Apart from genre related exploration: sum of IMDB Scores colored and sorted by keywords of a movie plot; most popular movies are based on books or novels. This query filters out results that were Null.



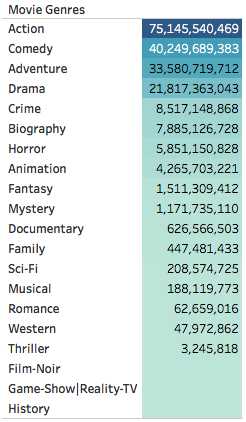
The bottom-tail of data for the same query as above:



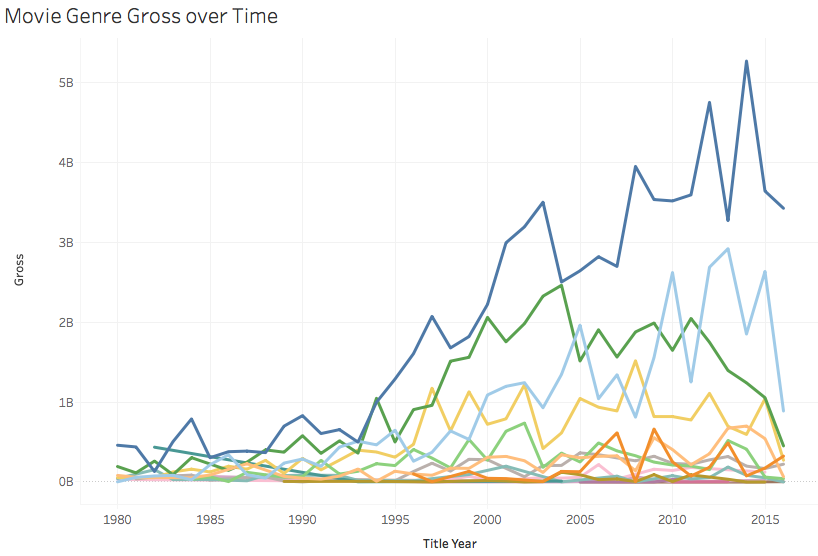
Back to genres. A comparison of movie genres by duration shows that comedy movies are typically the longest, while strictly historical films are the shortest in duration. The x-scale is in minutes.



Action and Comedy movies seem to make the most money, as measured by gross revenue generated over opening weekend. This explains the increase in their respective budgets over time.



This plot may as well be a self-fulfilling prophecy: if Action movies are believed to make the most, they will be allocated the most budgets, and thus also make more money over time.



Based on this EDA it will be interesting to explore predictive algorithms that can determine a movie genre by either duration or budget, or a combination of the two.