

Dataset Chosen: movies.csv

Analytical Tasks Supported:

- **Identify** the actors with the highest popularity on Facebook, the most popular directors, the least popular directors, the most popular rating of movies, the most popular language movies are in, and the most popular country a movie is made by.
- **Distinguish** among how the different attributes affect the popularity of films on social media, and among the popularity of films based on countries and languages.
- **Categorize** the films based on directors, actors, languages, country, ratings, and more
- **Cluster** the movies by grouping the movies made by the same actors, the directors with a similar number of Facebook likes, and the movies with a similar number of Facebook likes.
- **Rank** the actors by popularity, budget, gross, number of user for reviews, and IMBD score, the rating of films by movies Facebook likes, and the most popular languages and countries of films by movies Facebook likes
- **Compare** among the director Facebook likes, number of voted users, number of critics for reviews, number of users for reviews, movie Facebook likes, cast total Facebook likes, and IMDB score; among movie facebook likes, budget, gross, number of user for reviews, and IMDB score by actors; among different movie ratings; among different languages and countries;
- **Associate** the charts provided in pairs and triplets with each other.
- **Correlate** based on the scatterplots, bar charts, and packed bubble charts.
- **Retrieve** the year a film is produced when brushing and linking based on director Facebook likes, number of voted users, number of critics for reviews, number of users for reviews, movie Facebook likes, cast total Facebook likes, and IMDB score; the highest-ranking actors based on movie Facebook likes, gross, IMBD score, number of user for reviews, and budget; the name of the director when hovering over the scatterplot pairs; the percentage of each language and country when hovering over the donut pie chart.
- **Filter** languages and countries being compared by toggling the legend on the pie chart.
- **Compute the derived value** of the sum of the percentage of a certain number of languages or countries out of all movies.
- **Find extremum** of the year with the most film produced, the actor with the highest popularity of films, the actor with highest gross, the actor with the highest IMDB score, the actor with the highest number of user for reviews, the actor with the highest budget, the director with the most Facebook likes or movie Facebook likes, the movie rating with the highest and lowest popularity, the most popular language and most popular country in films.
- **Determine the range** of all attributes presented.

- **Find anomalies** in each set of visualizations.

Design Overview:

This visualization takes a “Storytelling” approach in answering the question— what does the number of likes a movie receives on Facebook imply about the movie? The page then takes the viewer on a journey to discover the answer to this question through a series of data visualizations. Each set of data visualization is presented with different attributes and accompanied by short blocks of text to help communicate the information presented.

The vis first presents the audience with a scatterplot and bar chart pair to give an initial impression of how the number of Facebook likes a movie receives result in other ordinal aspects of a film. To make it easier for the audience to explore the dataset at a glance, this first set of vis also supports brushing and linking. This helps the audience to understand the correlation among many of the entries provided by the dataset.

After gaining a general understanding of the dataset, the webpage then takes the user on a deep dive into how actors influence the popularity of the films. To make it visually appealing and easy to comprehend, a series of bubble charts were presented. This helps our audience to collect immediate takeaways simply by glancing at the size of each bubble presented. If desired, the user can also take the ordinal values provided by each bubble and perform computations to come up with new ideas and insights.

The scatterplot pair of the Director Facebook Likes vs. Movie Facebook Likes is displayed below the bubble charts. This set of vis supports tooltip, which is a functionality that supports hovering to see the names of the directors. Since the scatterplots only show ordinal data on the axis, having the nominal data (the names of the directors) show up on mouse over provides the audience with the details on demand interactions.

The next visualization is a bar chart demonstrating how ratings influence the movie Facebook likes. This chart uses two different colors to voice the message that movies rated R has the most Facebook likes.

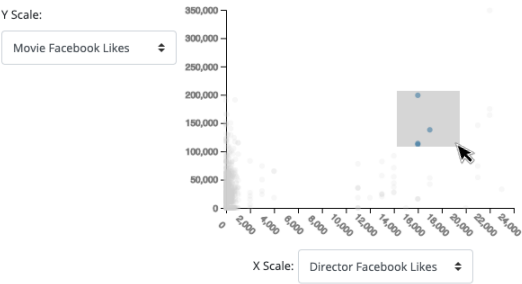
Lastly, the donut pie chart pair is then placed. This chart provides the user with an opportunity to interact with the data and observe new information outside of what is provided in the texts. By selecting and deselecting different languages or countries, the audience can filter in and out different attributes and establish new correlations based on countries and languages with the movie Facebook likes. This chart also provides details on-demand functionality— by hovering over the bands of the donut pie charts, the user can have a clear view of the percentage and the attribute each band represents.

This visualization aims to provide the user with ample information relating to how the different aspects of a movie would carry out a popular movie on social media. It uses multiple views, vis in pairs, and smooth interactions to help the user understand and learn about the popularity of movies led by different features. With the hovering, selecting, and toggling functionality provided in some of the visualizations, the user is also able to discover new information themselves.

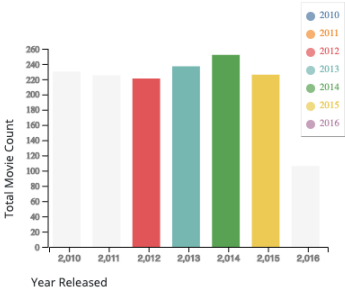
Vis Snapshots: (attached on the next page)

Vis #1: Brushing and Linking

Scatterplot for Exploration Purposes



Movie Released per Year



What Makes a Film Popular on Social Media?

The film industry has flourished in the past decade. With the increasing usage of technologies and innovations of streaming services, movies have become more accessible. Today, we will be examining what affects the popularity of movies with regards to their social media attraction— more specifically, **what does the number of likes a movie receives on Facebook imply about the movie?**

The dataset used in this visualization contains a collection of features for the films produced in the 2010-2016 timeline. (Note: the data containing “NaN” and “Null” entries were emitted during the visualization construction.)

To start, use the **scatterplot** and **bar chart** pair on the left to explore how the audience interacts with a film and the correlation among them through **brushing and linking**. While it may be difficult to observe a certain pattern, this visualization gives us an initial impression of some of the features that are contained in the dataset.

Vis #2: Packed Bubbles

The popularity of the actors have a significant impact on the popularity of the movies.

In the following **packed bubbles view**, we explore the actors' impact the number of movie Facebook likes. By summing up the total number of likes of all movies done by each actor, we observe the appearance of some actors with distinctively larger bubbles than the rest of the actors.

On the right, a zoomed-in version of the left bubble chart is displayed, giving us a closer insight on who are the top 34 actors with the most amount of movie likes. Needless to say, the top actors are all well-known, thus, this explains how the popularity of the actors impacts the popularity of movies.

Leading Actor Name vs. Movie Facebook Likes



Leading Actor Name vs. Movie Facebook Likes (Top 34 Actors)



This phenomenon of popular actors resullt in more social media interactions for the movies can also be reflected by the gross.

In this second pair of the bubble chart, we further examine how much can an actor really impact the popularity of the movie by looking at the total gross of the movies lead by each actor, and we observe that familiar names, such as **Jennifer Lawrence** and **Chris Hemsworth** from the previous bubble charts stand out in this set of bubble charts as well. This observation once again proves that the more popular the actors are, the more likely people are willing to watch the movie.

Leading Actor Name vs.Gross



Leading Actor Name vs. Gross (Top 32 Actors)



The popularity of actors also impacts the score of the film, and the more popular the actors are on social media, the higher the budget for the film itself.

In the following three bubble charts, we attempt to find a correlation between each movie actor and the total IMDB score received by the movies, the total number of users for reviews of the movies, and the budget for the total number of films acted by the actor. Due to a large number of entries in the dataset, a lower bound was set for each feature for us to more efficiently observe the resulting bubble views.

Leading Actor Name vs. IMDB Score



Leading Actor Name vs. Num User For Reviews



Leading Actor Name vs. Budget

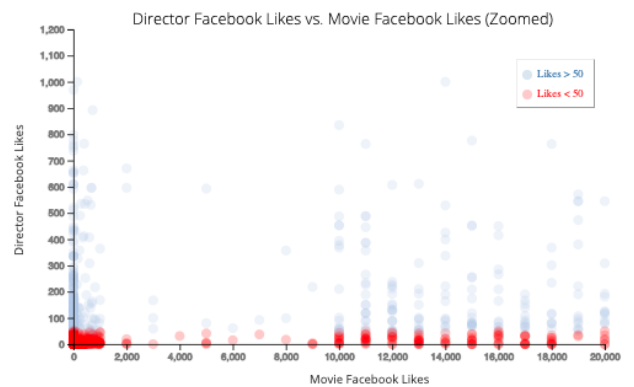
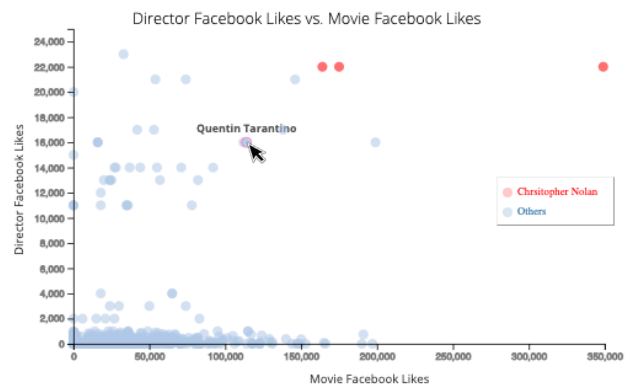


Vis #3: Scatterplot with tooltip

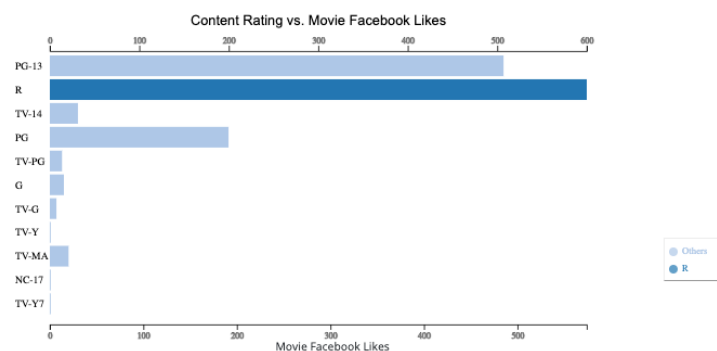
The popularity of directors is somewhat positively correlated to the popularity of the films.

Take a look at the **scatterplot** on the right, a positive correlation can be established by only looking at the directors with Facebook likes of more than 10,000. Below such cutoff, we can rarely see a correlation between a director's Facebook likes and movie Facebook likes. The dots highlighted red represents the director, Christopher Nolan, who has the most overall Facebook likes, and we can see that his films have performed significantly better than all the others when received by social media. If you **hover** over each individual dot, you can see which director it represents.

Let's take a look at the zoomed-in version of this lower half of the scatterplot, we see that there is neither a positive correlation nor a negative time correlation between the director's Facebook likes and movie Facebook likes. Especially the regions highlighted in red, we see how even the directors receive lower than 50 likes, the performance of their films is not relative to themselves.



Vis #4: Bar chart



Films rated R gain the highest popularity on social media.

The **bar chart** on the left explores how the ratings of the film influence the popularity of social media. We see that films rated R has the highest overall movie Facebook likes.

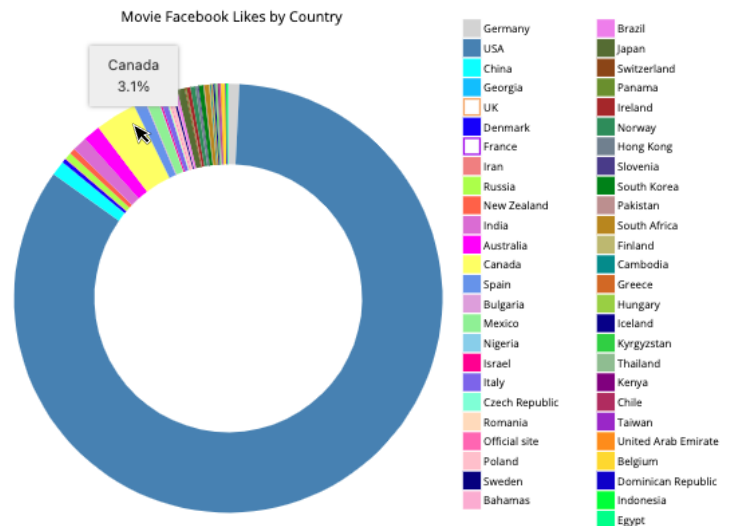
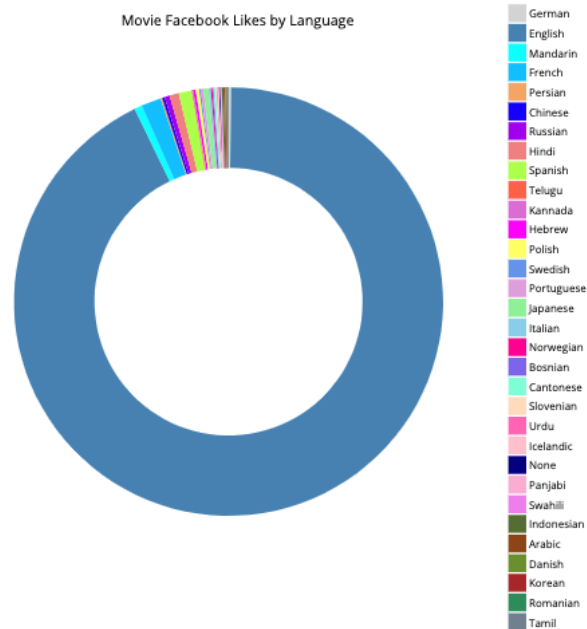
Vis #5: Pie charts (with filter & tooltip)

Movies in English and made in the USA have the highest overall popularity on social media.

Finally, the two **donut pie charts** on the right explore the movie Facebook likes based on the language and country of the film. We observe that English films and moves made in the USA have the overall highest number of likes on social media.

Toggle the squares on the legend to **filter** out certain languages and countries and make observations based on the resulting pie charts.

Note: If you deselect English and USA, you can see a drastic change in the ratios among different languages and countries and make new observations based on the resulting pie chart. You can also **hover** to see what percentage each attribute obtains. **Try it out yourself!**



Example of filtering:

