

“Movie Madness” Process Book

CS 5630

Kaden Hendrickson, u1200510@utah.edu, u1200510

Ethan Ramos, u1182994@utah.edu, u1182994

Project Repository:

<https://github.com/ethramos22/dataviscourse-pr-moviemadness>

Table of Contents

“Movie Madness” Process Book	1
Table of Contents	2
Background and Motivation	3
Project Objectives	3
Data	3
Visualization Design (Kaden & Ethan)	4
Design Evolution	10
Movie Genre Color Scheme	10
Displaying revenue in the overview table	11
Implementation	12

Background and Motivation

Growing up we loved watching movies, and have also struggled trying to find what movie to choose. There are so many options in today's world with streaming platforms as well as cable and weekly releases at the theater. We want to stay in the know of pop culture by focusing on what is popular, relevant, and rated well to help us and others decide what to watch.

Project Objectives

Benefits:

- Help users determine what show or movie to watch, by filtering data by genre, popularity, score, platforms, and more

Accomplish:

- See trends amongst popular and highly rated movies based on budget, genre, revenue, and more
- Point users to streaming platforms and places that currently offer a selected movie
- Show movies profits and losses based on budget and revenue
- Show which countries had produced a selected a movie

Answer:

- What movie is the most popular and has the highest reviews and vice versa?
- What movie is the most popular with the lowest reviews and vice versa?

Data

We'll be collecting our data from 'The Movie Database' aka TMDB (<https://developers.themoviedb.org/3/getting-started>). This movie database is free to use, and contains massive amounts of data related to movies, tv shows, streaming platforms, actors, geographical location, etc. We'll utilize multiple endpoints, the most relevant being the following:

- /movie/{movie_id}
- /movie/{movie_id}/images
- /movie/{movie_id}/similar
- /movie/popular
- /movie/top Rated

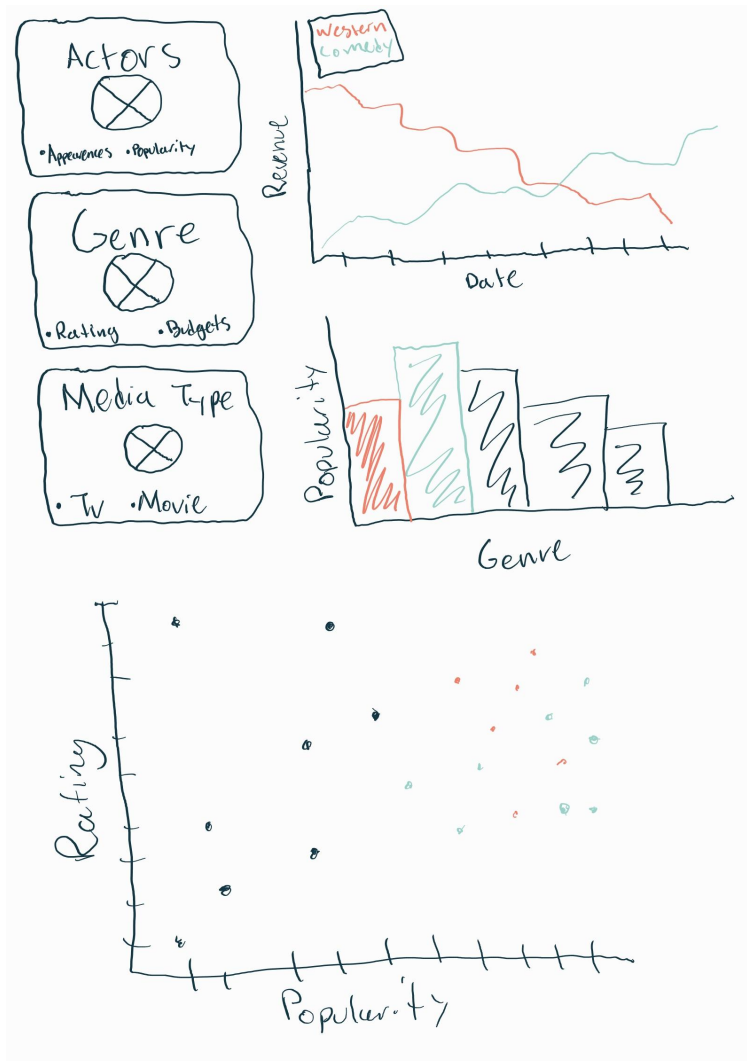
Using TMDB will require minimum data cleanup with its clean and easy to use API. Our main derivation from the TMDB API will be for a movie Id. One of our main API calls will be a movie title. From the result we can take the "id" field to make additional queries on the details of that movie. This will lead us to the details API call for a movie. From this response we can retrieve things like genre, budget, revenue, poster url for an image of the poster, popularity rating, and runtime.

Another main query for our project will be "Get Popular" to retrieve an array of the most popular movies. This query can return an array of the movies most popular by view count, and is updated on a daily basis. Similarly we will use "Get Top Rated" and "Get Upcoming" to populate our movie list with relevant data.

For our data processing we can use the returned json objects to bind to d3 elements and javascript objects. Packaging the movie into its own object for later use on multiple visualizations.

Visualization Design (Kaden & Ethan)

- Interactive feature (in the shape of a card) on the left (choose genre, actor, industry, media type, etc.) on the right we get interactive visualizations related to the feature you chose (profit, budget, rating, popular actors)



Movies

☒ Title
.Profit . Genre

☒ Title
.Profit . Genre

☒ Title
.Profit . Genre

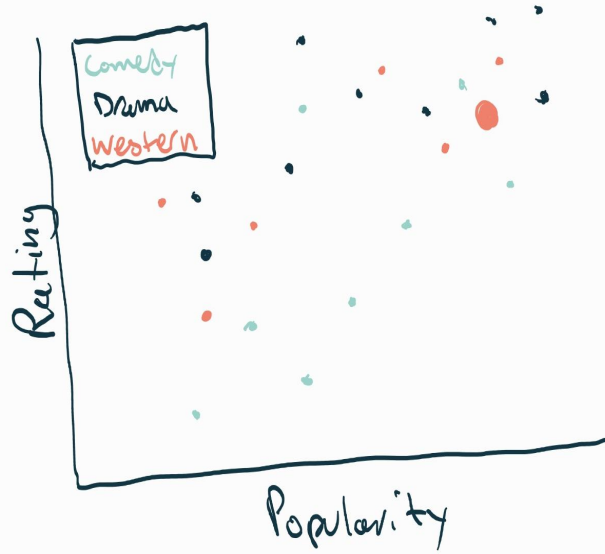
☒ Title
.Profit . Genre

☒ Title
.Profit . Genre

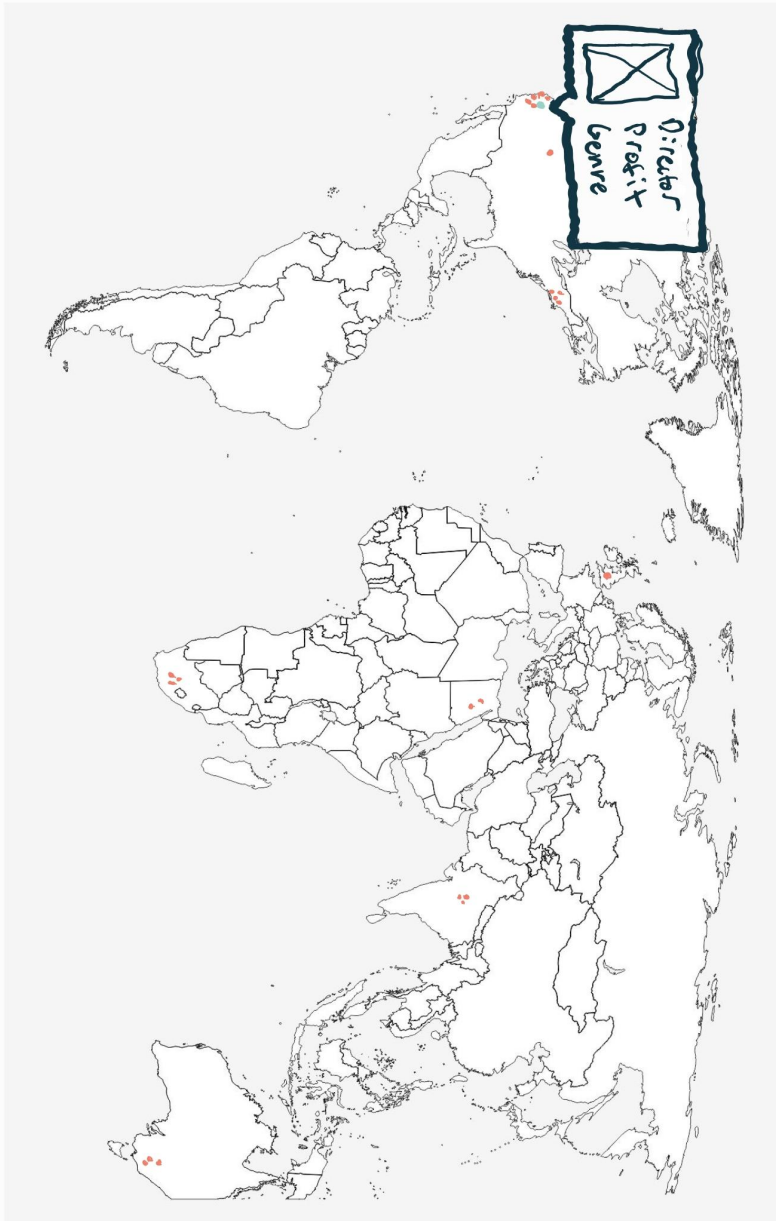
☒ Title
.Profit . Genre

X-axis ▼

Y-axis ▼

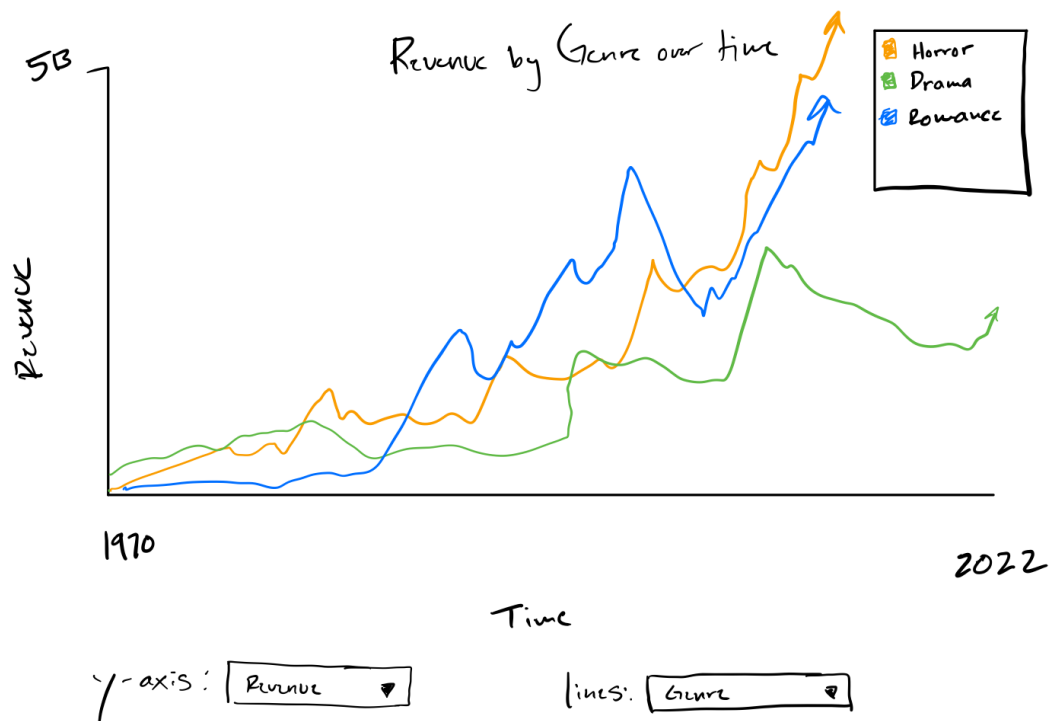


- Map with pins on production location of popular movies. When a location gets clicked, a view appears with more detailed information about the movie/movies that were produced in that country.

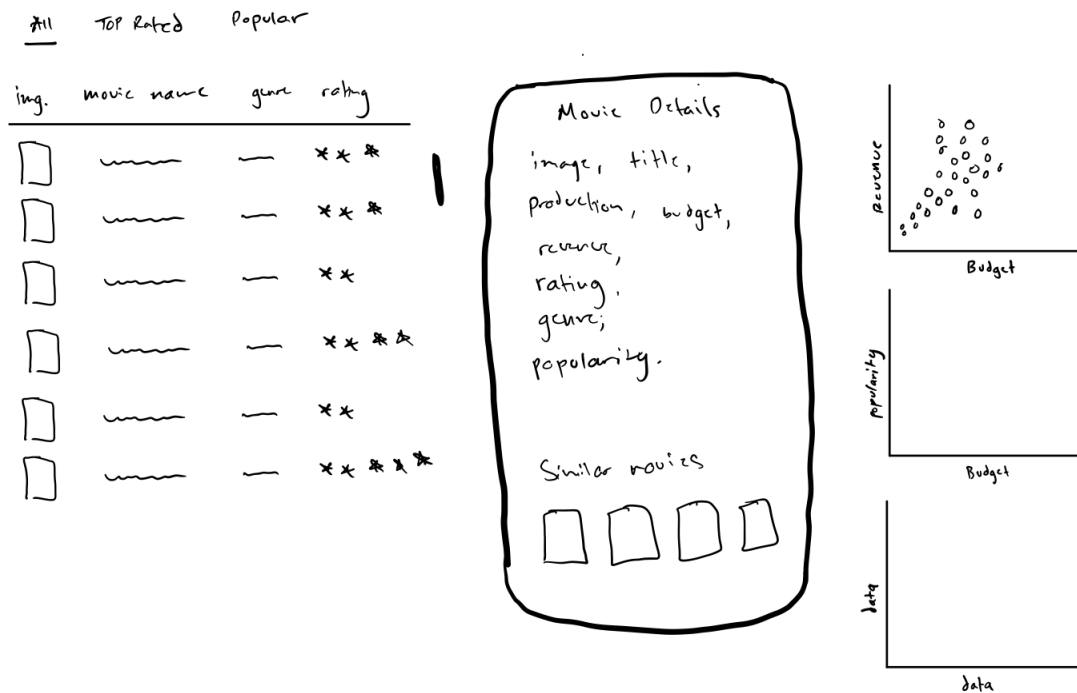


Movies by production location

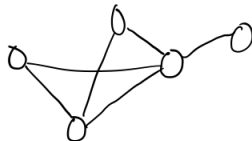
- Line chart that maps quantitative movie data over time for different genres (ie. change in budget/revenue/production time/ etc. over time for drama movies). This visualization could allow changing in the y axis or in the data show by manipulating a dropdown menu



- Final Design:
 - Will contain a list of movies that can be filtered by popularity, top rated, upcoming, etc. The list will contain visualizations about the movie info (rating, genre, language, adult content, etc.). A panel in the middle of the screen will display the movie, and all the details we can find about it. The panel on the right of the screen will display several visualizations, and will highlight the currently selected movie in each of the visualizations - to show it compares to the other movies in the list. Potentially, we can create another visualization on the bottom of the screen that will show geographical or timeline data. Maybe even actors being connected to one another.



Actor Connectedness



Exploratory Data Analysis

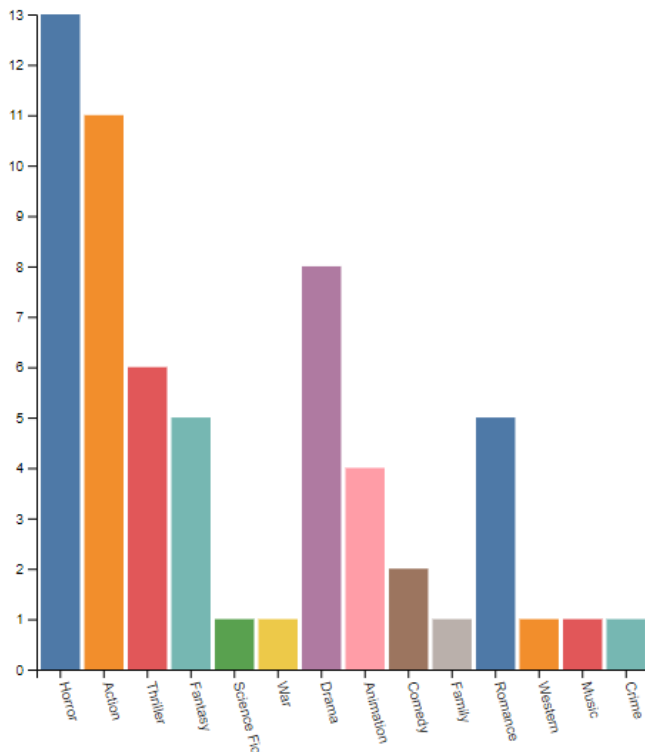
Our initial exploration of the data solely involved parsing through the json to compare attributes and see what kinds of visualizations we could create. Once we had a general idea of the data we were going to use, we used scatterplots to plot different attributes against each other. After testing out plotting multiple combinations of attributes, we were able to find which combinations were most insightful to the user (Budgets vs. Revenue, Budget vs. Rating, and Revenue vs. Rating)

Design Evolution

Distribution Chart Addition

[entry by Ethan, 11/7]

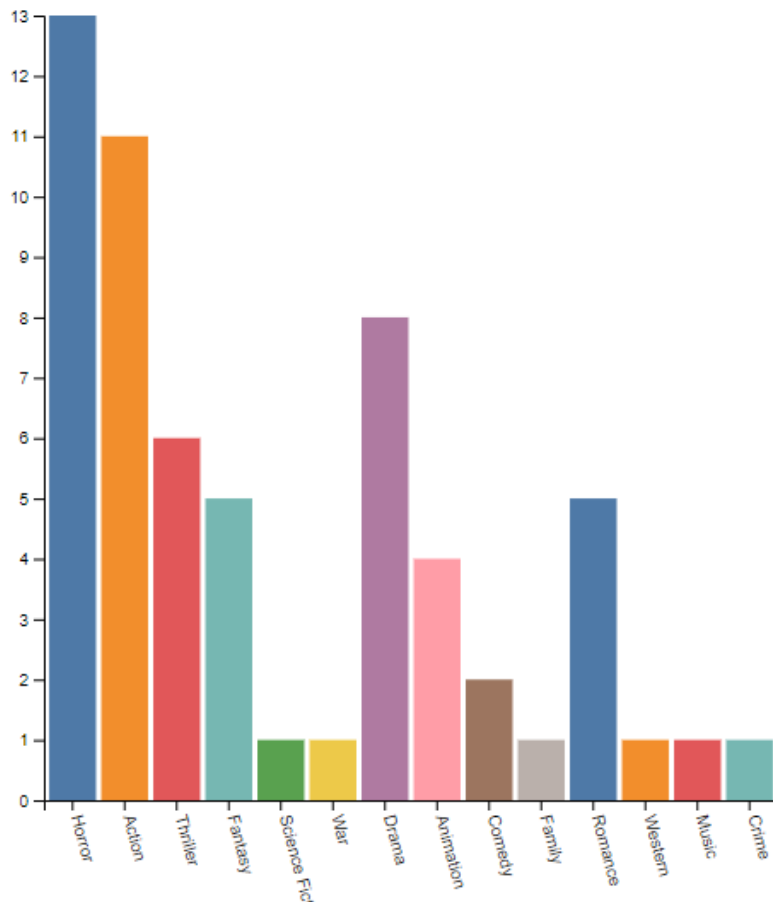
Based on feedback from Haihan Lin, an overall distribution chart will be necessary for our project. Not only will this add more visualization, but also show overview data of all the selected movies. Firstly we can show how many movies have a specific genre. The current dataset based on most popular movies, currently playing movies, and top rated movies. Data was collected the week of halloween and we can see that horror is the top genre.

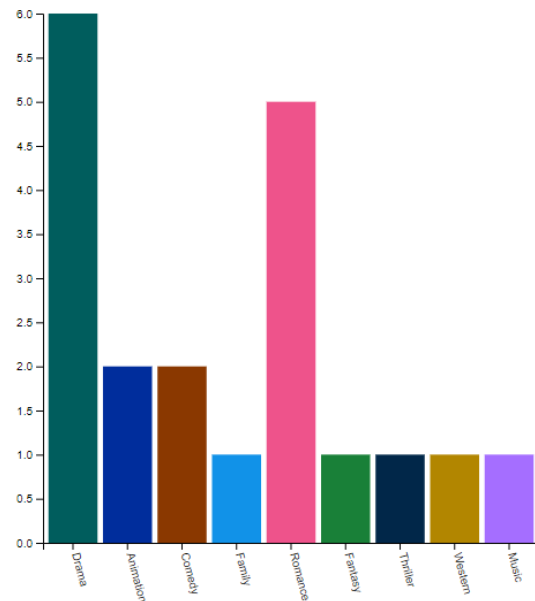
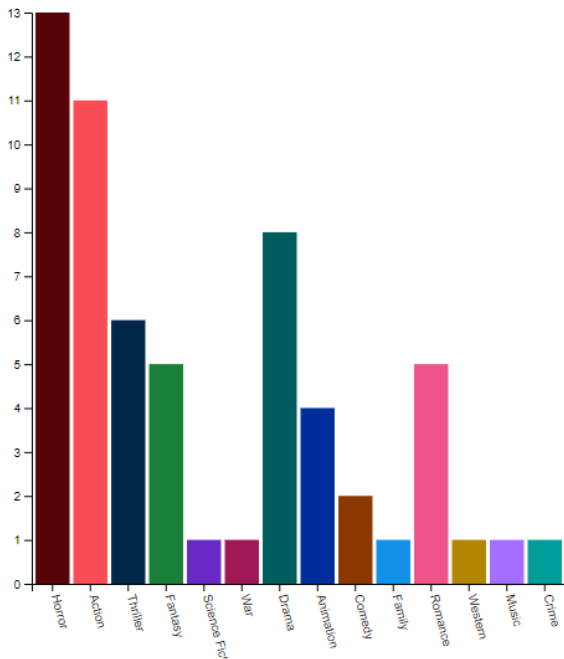


Movie Genre Color Scheme

[entry by Ethan, 11/9/22]

Using the d3 set color scheme we would get duplicate colors for different genres, due to the limitations of using a preset scheme as tableau10 would only have colors for 10 values. We knew our data needed an ordinal color scale with a value for each genre. Color coding genre's will help with our other visualizations on our web page, for example to color code the dots on our dot plots.





Using [color palettes](#), we could get an x amount of color values for an ordinal scale. From our data we had 13 genres. We created our range of these 13 color values to map 1-to-1 to the keys of the data grouped by genre. When the data would change from selection and interaction, genres will always have the same color. As we can see above, changing the list of movies from popular to top rated, the Drama genre becomes the top genre.

Displaying revenue in the overview table

[entry by Kaden, 11/9/22]

Initially, we thought that revenue would be a good metric to list in the table, to give viewers an 'at-a-glance' understanding of how much money the movie generated. We quickly found out however that the range of values was \$0 to \$3 billion, which is a hard range to visualize inside an 80 pixel svg. We decided it would make more sense to display the movie runtime - as it is a more digestible metric, and is possibly a more valuable value for the user to get 'at-a-glance.'

Movie Name	Genre	Rating	Language	Revenue
Terrifier 2	Horror, Thriller	71%	English	\$5,325,078

TODO: Screenshot of runtime column

Implementation

Describe the intent and functionality of the interactive visualizations you implemented. Provide clear and well-referenced images showing the key design and interaction elements.


Table item selection

[Kaden, 11/9/22]

Clicking on an item in the tables displays more information about the selected movie in the detail card to the right of the screen. This allows us to show more information than what the table can hold, including the movie poster, movie description, and eventually a list of 'similar movies'.

Movie Name	Genre	Rating	Language	Revenue
Terrifier 2	Horror, Thriller	71%	English	\$5,325,078
Black Adam	Action, Fantasy, Science Fiction	68%	English	\$19,000,000
Blade of the 47 Ronin	Action, Fantasy	66%	English	\$0
Terrifier	Horror, Thriller	66%	English	\$0
MexZombies	Horror, Action, Comedy	74%	Spanish	\$0
Fall	Thriller	73%	English	\$11,900,000
Jeepers Creepers: Reborn	Horror, Thriller, Mystery	57%	English	\$2,892,594

Blade of the 47 Ronin
106 minutes Action, Fantasy



In this sequel to "47 Ronin," a new class of warriors emerges among the Samurai clans to keep a sought-after sword from falling into the wrong hands.

Movie collection selection

[entry by Kaden, 11/9/22]

Users are able to toggle between different collections of movies, including 'All', 'Now Playing', 'Popular', and 'Top Rated'. Toggling between these selections also changes the distribution chart and the smaller dot-plots at the bottom of the screen, to display data relevant to the selected movie collection.

All

Now Playing

Popular

Top Rated

Movie Name	Genre	Rating	Language	Revenue
				<div> <div></div> <div>10</div> <div>\$2,000,000,000</div> <div></div> </div>
The Godfather	Drama, Crime	87%	English	\$245,066,411
The Shawshank Redemption	Drama, Crime	87%	English	\$28,341,469
The Legend of Hei	Animation, Fantasy, Action	87%	Chinese	\$48,045,728
The Godfather Part II	Drama, Crime	86%	English	\$102,600,000
Schindler's List	Drama, History, War	86%	English	\$321,365,567
Dilwale Dulhania Le Jayenge	Comedy, Drama, Romance	86%	Hindi	\$100,000,000
Impossible Things	Family, Drama	86%	Spanish	\$0