

# Project Proposal

## Basic Info

- Project title: Moviz
- Benjamin Mastripolito [u1419419] ([benpm@cs.utah.edu](mailto:benpm@cs.utah.edu))
- Alper Sahistan [u1406260] ([alper@cs.utah.edu](mailto:alper@cs.utah.edu))
- <https://github.com/benpm/moviz>

## Background & Motivation

We like stories that individual movies tell, but we also think movies collectively tell us stories. Visualizing the movie data, such as budget, rating, release date, oscar, and genre, can show us interesting information about individual movies, trends, industry direction, and much more. We also want a practical tool that we can use to look up movies to decide what to watch that day.

We think that rather than letting user search through uninteresting visualizations, we could tell several interesting stories we pre-defined while giving room to explore meaningfully. Therefore we decided to lead user through curated three different views which can be switched between.

## Project Objectives

There are multiple things we are trying to convey through our visualization:

- What film genres compose oscar nominations and winners over its history?
- How can we display film data for both film discovery and for learning interesting trends from film history?
- How has the market share of various film studios changed over history?
- What benefit is gained to a film's popularity, commercial success, and rating by having a larger budget?
- How have different measures of success, measured together, changed for films over the decades?
- What relationship do measures of financial success and measures of popularity share among films over the years?

## Data

- (Oscar data set) <https://www.kaggle.com/datasets/unanimad/the-oscar-award>  
Associating film titles from the Movie Industry data set with the following
  - Oscar Nominations

- Oscar Wins
  - Oscar Best Picture Awards
  - Oscar Year
- (Movie Industry data set) <https://www.kaggle.com/danielgrijalvas/movies>  
 This will be our primary data source, and we will only be using films in this data set
  - Film Title
  - Profit
  - Production Company
  - Runtime
  - Directory
  - Budget
  - Release Date
  - Genre
  - IMDb User Rating
  - IMDb User Votes
- (IMDb data set) <https://www.imdb.com/interfaces/>  
 We may use this for additional information such as multiple genres and language

## Data Processing

The only data processing we plan to perform is to associate film titles from the Movie Industry data set with Oscar information from the Oscar data set. We will achieve this by writing a simple Python script to associate the data and output it into a final, cleaned CSV.

We might have to do some other data aggregation for heatmaps and trendlines yet those are to be seen during the implementation.

## Visualization Design

Since we want to visualize meaningful channel pairings we only allowed certain combinations. Our data will be displayed via a [data-collection dashboard](#) user interface with multiple plots and options for modification.

We brainstormed several ideas to derive three different plot views.

## Brainstorming Sketches

## ## Visualization for Data Science - Vis ideas ##

### Data

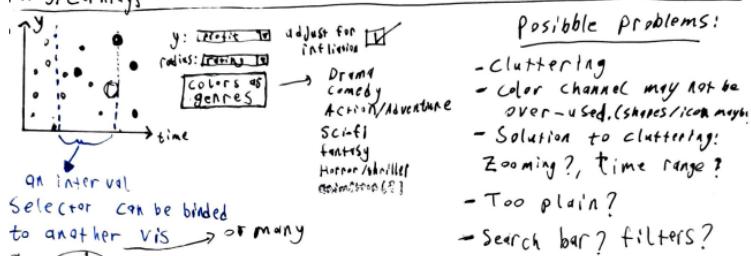
#### Plain

Name, date, rating, genre, cost, earnings, studio(?)

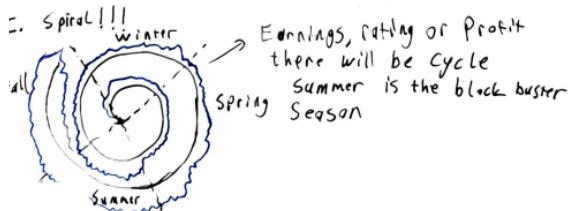
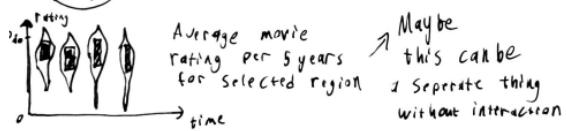
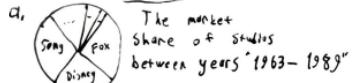
#### Derived

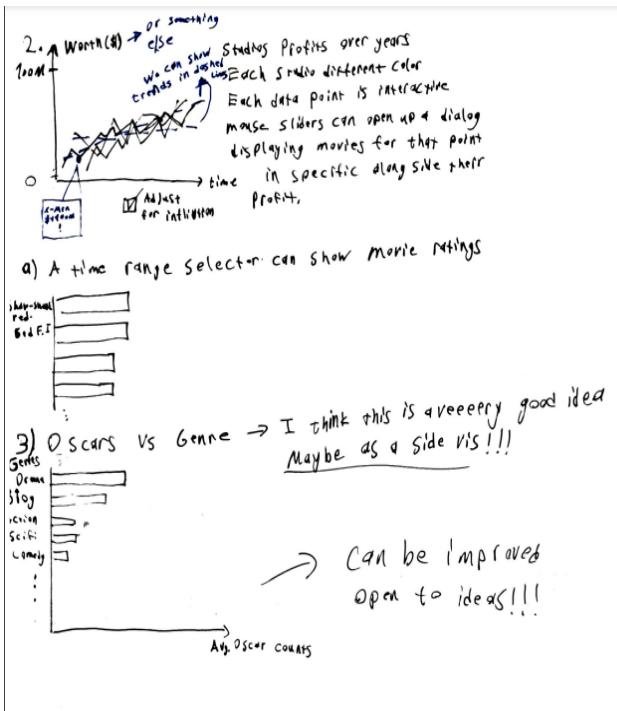
Profit, inflation adjusted; cost, earnings, profit, rating/cost, rating/profit,

#### Rating/earnings



Selector can be binded to another vis



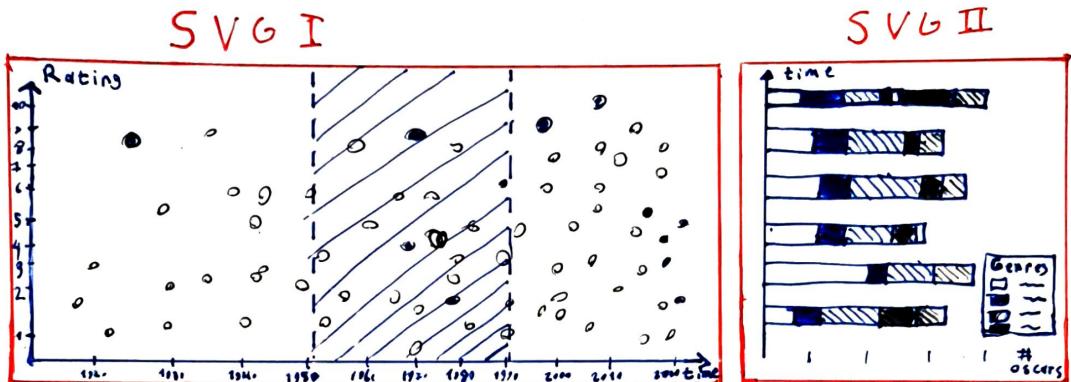


1. Movie Discovery View
- a) Scatter Plot - Main  
 $y = \text{ratings}$   
 $x = \text{time}$   
 $\text{area} = ?$   
 $\text{Color} = 4 \text{ colors} \rightarrow$  nominated  
 won  
 nominees best pic.  
 won best pic.
- b) Stacked line chart - Sub  
 $y = \text{Oscar nominations, Oscar winners}$   
 $x = \text{time (Selected from main)}$   
 $\text{mark} = \text{Lines (Stacked)} \rightarrow \text{Genres}$   
 $\text{Color} = ?$
2. Budget & Studio View
- a) Scatter Plot - Main  
 $y = \text{Profit, cost, earnings}$  → Line (Avg.)  
 $x = \text{time}$   
 $\text{Color} = \text{green/red (redundant)}$
- b) Pie chart - Sub  
 $\text{angle} = \text{avg. profit for selected period of time (Market Share)}$   
 $\text{Colors, text} = \text{studios}$
- 3.
- 

## Prototypes

These prototypes are the ideas we think should make to the final design since they each tell a different but meaningful story. They will be combined into a final dashboard, and the dashboard will let users switch between these visualization pairs.

## Design A: Movie Quality Exploration: Rating vs. Time & Oscar Distribution Over Selected Time Interval



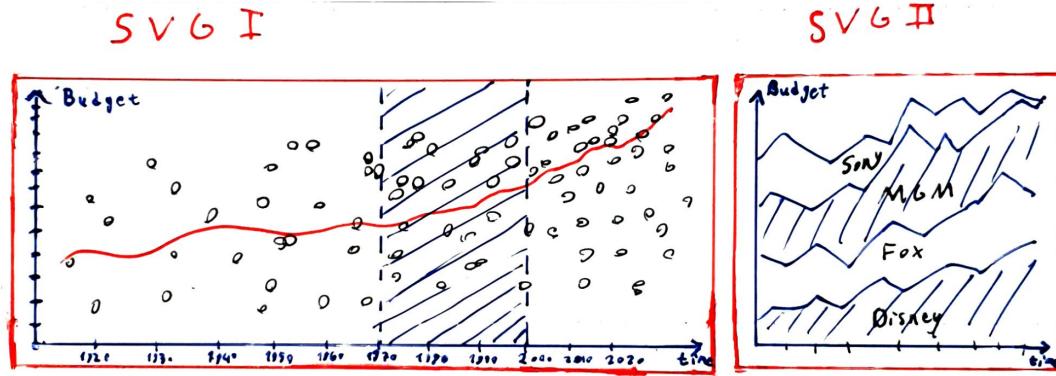
X axis	Release date
Y axis	Rating from IMDB or Rotten Tomatoes
circles	Movies
colors	Oscar Nominee #14591D Oscar Winner #99AA38 Oscar Best Picture Nominee #5C2751 Oscar Best Picture Winner #258EA6

X axis	# Oscars
Y axis	Selected Years
Stacked Bars	# Oscars for the genres
colors	Genres

In this design, we have a scatterplot in SVG I with the X axis as the release date and the Y axis being the rating(IMDB, Rotten Tomatoes, etc.). Points on the plot will also be colored by their Oscar information. Different colors will be associated with nominated films, winners, and best picture nominees and winners. The user can hover to reveal a tooltip showing specific film information such as title, director, studio, budget, and profit. Moreover, the user can zoom, pan to interact with the scatter plot. Another way to interact is by selecting a time range on the scatterplot, which will also change the time range on SVG II. The visualization on SVG II will

display a stacked bar chart of genre distributions for Oscars for selected years. The data in SVG II can be bound to either nominations or winners.

### Design B: Economy of Movies: Profit vs. Time & Studio Market Shares Over Selected Time Interval



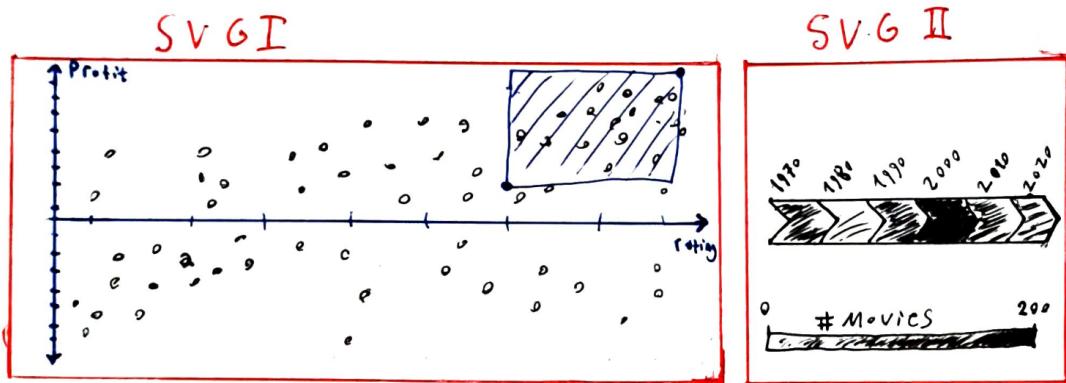
X axis	Release date
Y axis	Budget, Cost or Profit
circles	Movies
colors	Emphasize profit or loss again

X axis	time
Y axis	Spending, Profit, Market share
Stacked Areas	Studios

In this design, for SVG I, we have a time X axis like Design A, but with a budget, cost or profit Y axis. The secondary visualization on SVG II shows a stacked area chart visualizing the market share between major production companies such as Sony, MGM, Fox, Disney, etc.

To illustrate average budget, cost or profit, user can toggle to show the average across all films as a line overlaid atop the scatterplot. We also color the points based on whether they had negative or positive gross earnings. Again, the user can zoom, pan, and select individual points to see a tooltip with more detailed information.

### Design C: Movie Cost Effectiveness: Ratings vs. Profit & Per-Decade Heatmap for the Movies in the Selected Region

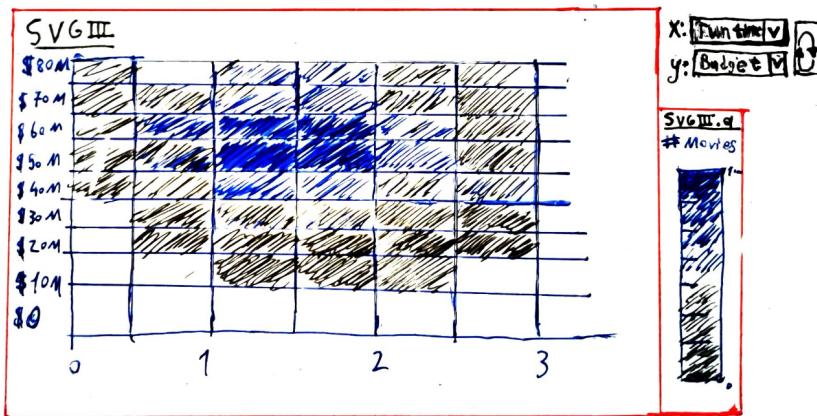


X axis	Rating: IMDB or Rotten Tomatoes
Y axis	Budget, Cost or Profit
circles	Movies
colors	Emphasize profit or loss again

X axis	Decade bins
Color	# Movies

In this design, for SVG I, we have a similar layout to designs A and B; however, we remove the time axis to combine profit metrics with quality ratings on a scatter plot. The Y axis can be bound to the cost and budget of the given movie, and the X axis can be bound to different quality ratings. Additionally, a selection on the scatterplot can be made by selecting a rectangle area which changes the display for SVG II. The visualization on the right side has a per-decade heatmap where the values within the cells represent the number of movies in the selected area in the scatterplot.

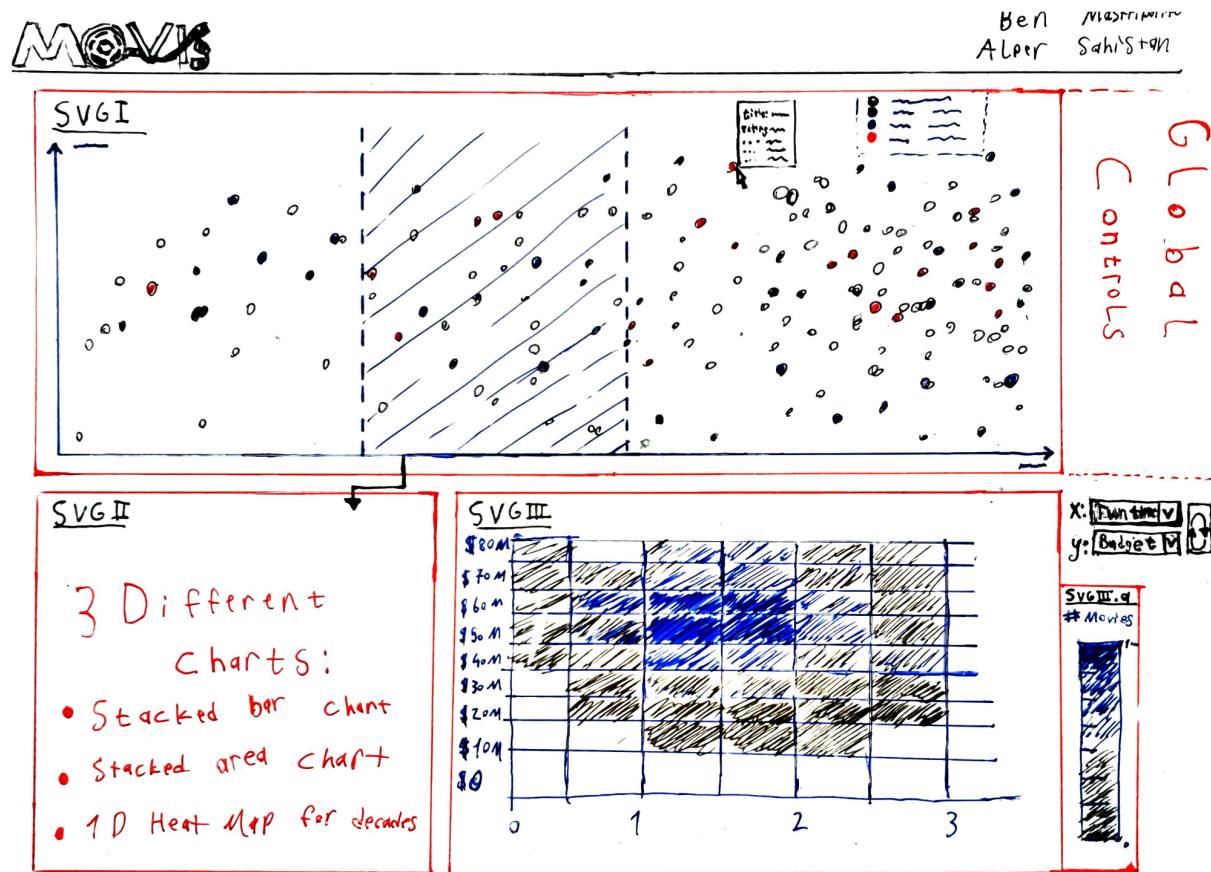
## Design D: Movie Distribution Explorer: 2D Heatmap with Pickable X and Y axes



X axis	Free
Y axis	Free
colors	# Movies

This design aims to let user explore several metrics freely without clutter as it is a density heatmap. Unlike previous designs this design is bound to a separate SVG element (SVG III). Therefore it will always be present. The quantities can be chosen by the user from the set of budget, profit, cost, rating, release date and # oscar nominations.

## Dashboard Design



The dashboard design incorporates the previous three designs bound to SVG I and II as well as the extra 2D heatmap design bound to SVG III. Depending on what the user picks via "Global Controls" the whole dashboard will change views. Views are the prototype designs A, B and C. SVG III will let user explore certain aspects of the the data however whenever the view changes we will re initialize the axes to values we think compliments the selected view (We may remove this feature if it becomes disturbing).

## Features

### Primary

- Three separate views which can be selected by the user
  - View one: release date / rating
    - Secondary visualization: stacked bar chart of distribution of Genres in Oscar nominations
  - View two: release date / budget, profit, cost

- Secondary visualization: stacked area chart of market share of various production companies. Can be shown as market share in dollars or percentage
  - View three: rating / budget, profit, cost
    - Secondary visualization: quantity of selected films per decade
- Selectable axes for the scatterplot
  - View two: budget, profit, or cost
  - View three: budget, profit, or cost
- Heat map visualization
  - Selectable axes: budget, profit, cost, rating, release date, run-time and # oscar nominations
  - Color legend
- Detail tooltip
  - Shown when selecting a point on the scatterplot

## Optional

- Data point coalescing in scatterplots to avoid overlapping
- Transition animations when swapping between views
- Tooltips for the heat map showing details of selected cell

## Schedule

- **Week 1** (10/23 - 10/29): Beginning of basic layout, boilerplate
- **Week 2** (10/30 - 11/05): Basic layout and plot boilerplate
- **Week 3** (11/06 - 11/12): Scatterplot functionality
- **Week 4** (11/13 - 11/19): Heatmap functionality
- **Week 5** (11/20 - 11/26): Secondary plots functionality
- **Week 6** (11/27 - 12/02): Finishing touches and extra functionality