

# Project Proposal

## Basic Info:

<b>Project Title</b>	Movies Glorious Movies		
<b>Group Members</b>	Don Harrison (u1131452)	Sara Nurollahian (u1217653)	Julia Ma (u0918404)
<b>Email</b>	<a href="mailto:u1131452@utah.edu">u1131452@utah.edu</a>	<a href="mailto:u1217653@utah.edu">u1217653@utah.edu</a>	<a href="mailto:u0918404@utah.edu">u0918404@utah.edu</a>
<b>Repo:</b>	<a href="https://github.com/juliama920/dataviscourse-pr-Sajudo">https://github.com/juliama920/dataviscourse-pr-Sajudo</a>		

## Background and Motivation:

The rise and fall, as well as acquisition and merging of media companies is of particular interest to us. We all like movies and this dataset provides interesting features to explore, particularly budget, revenue earned by a movie, its company and the review score. We hope that visualizing the success of movies as well as information about the success of companies will give us a clue about why companies succeed, fail, get bought out, or merged.

We also are interested in seeing what drives a movie's popularity over time. Is it purely because of the budget spent on the film? Does a movie's success financially correlate with the reviews it gets online? These two questions are what motivated us to choose this project.

## Project Objectives:

As we all are interested to learn more about movies and their companies' success, we use this project as an opportunity to first understand which movies and their companies were successful over time and how different factors play a role in their success. Some of the questions we are interested to examine are:

- What genre was more successful at what time?
- What distributors were more successful?
- The relationship between reviews and the monetary success of movies?
- How did movie distribution companies grow over time?

## Data:

Our data is collected from Kaggle. A combination of the following two data sets will answer our questions relating to movies, release dates, distributors, viewer ratings, genres, gross income, and more.

"Highest Holywood Grossing Movies.csv" SANJEET SINGH NAIK: [Top 1000 Highest Grossing Movies | Kaggle](#)

"Movies(1986-2016).csv" DANIEL GRIJALVA: [Movie Industry | Kaggle](#)

We found an additional dataset that contains historical company stock information. We believe it will be interesting to compare the distributor stock information to its movie release information but may or may not use it. This data was also found on Kaggle.

"NASDAQ Historical Data" OLEH ONYSHCHAK: [NASDAQ dataset](#)

## Data Processing:

Most of the data we require is in the datasets we found on kaggle. As an optional feature, we were considering comparing the movie gross income to its distribution company's stock data. We would compare the success of a movie to its distributor's company success as a whole. We found a dataset that contains historical stock data of companies over a range of years. To get the stock data, we would need to collect all the movie distribution companies off our list of 1,000 highest grossing movies and retrieve their Nasdaq keys from the Nasdaq symbols metadata set. Then we could download the individual company stock csvs which contain the company stock data per year and compare them to our movie data. This would be the only data we would have to spend some time to acquire. Otherwise we do not expect to spend much time cleaning data.

Between the two movie datasets, we plan to use the following:

- Movie Name
- Distributor Name
- Original Release Date
- World Sales
- Main Genre
- IMDb Score

## **Must-Have Features:**

- **Line Chart (Company Success/Movie Chart)**
- **Bubble Chart (Movie Rating/Revenue Chart)**
- **Stream Chart (Revenue/Genre)**
  - These are the visualizations driving our data and without any one of these we feel the visualization will be lacking in painting a whole picture.
- **Tooltips on selection giving more info**
  - Since each movie contains so much info, we are using three separate but interconnected visualizations to portray our data. To help with the interconnectedness, we want mouseover and click event tooltips to appear on various elements on all the graphs. These tooltips will portray missing data on the visualization and connect the three graphs.
- **Click event on one graph updates every graph**
  - We want a selection of a movie on one graph to select the movie on all the graphs and highlight the movie across the whole page. Again, it connects the visualizations and allows viewers to see the movie's information (company success, rating, revenue, and genre) across the whole page.

## **Optional Features:**

- **Search bars on tables**
  - Since there are many companies to choose from on the Company Success/Movie Chart, we want to add a search bar to the graph so you can go ahead and search for the line to select on the chart.
- **Add transitions to dots on Company Success/Movie Chart**
  - When a company line is selected, dots which show the movies published by that company would grow on the line.
- **Table sorts on different selections.**
  - We want to add a table with additional information about the movies in our data set, such as director, country produced in, main actors, etc. We want this table to sort by certain fields based on what you've clicked on in the page. For example, if you select a company line on the Company Success/Movie Chart, it will sort the table by Company, so you can see the extraneous information about the movies produced by that company in the table.

## **Project Schedule:**

### **Oct 21: Project Proposal Due**

- Turn in project proposal and determine individual responsibilities

### **Oct 28:**

- Data ready to use
- Webpage set up with svgs holding our visualizations

### **Nov 4:**

- Individual graphs working but not interacting
- Process book set up

### **Nov 11: Project Milestone**

- Work towards class milestone submission
- Update process book
- Completed Data Acquisition
- All Data structures in place
- Working visualization prototype

### **Nov 18:**

- Start interaction
- Update process book

### **Nov 25:**

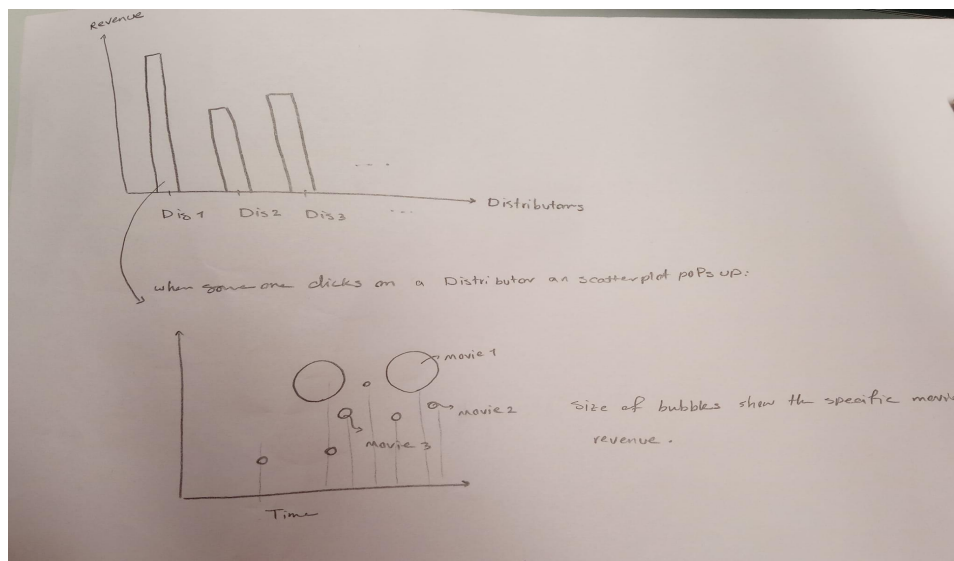
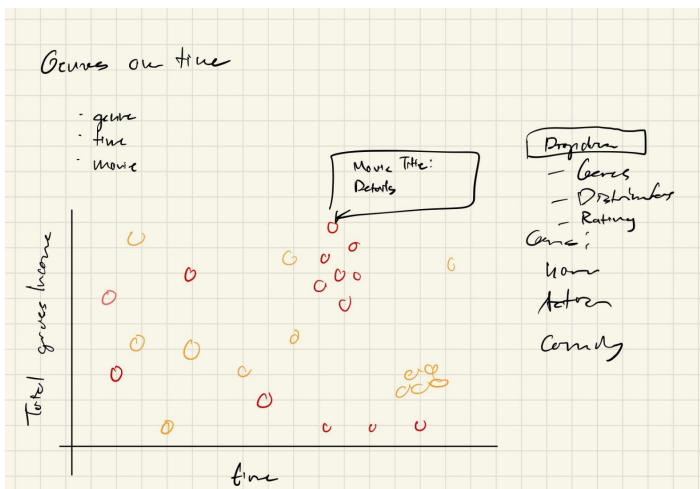
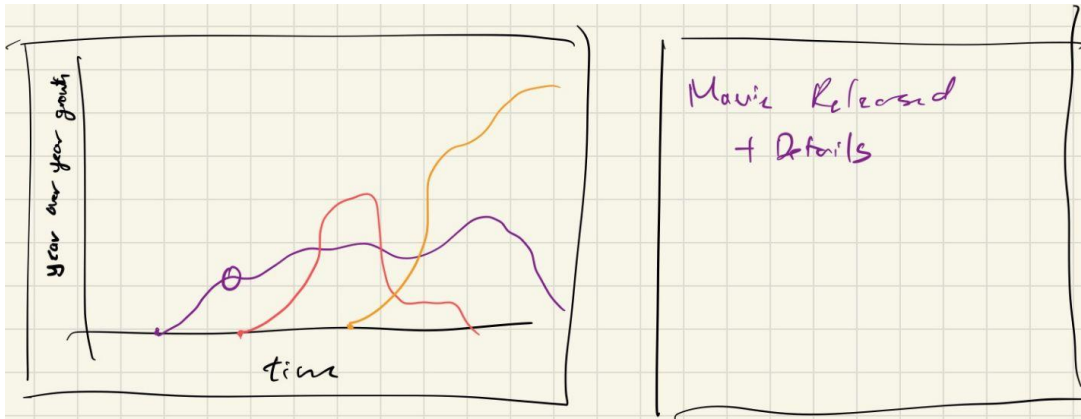
- All graphs interacting
- Update process book

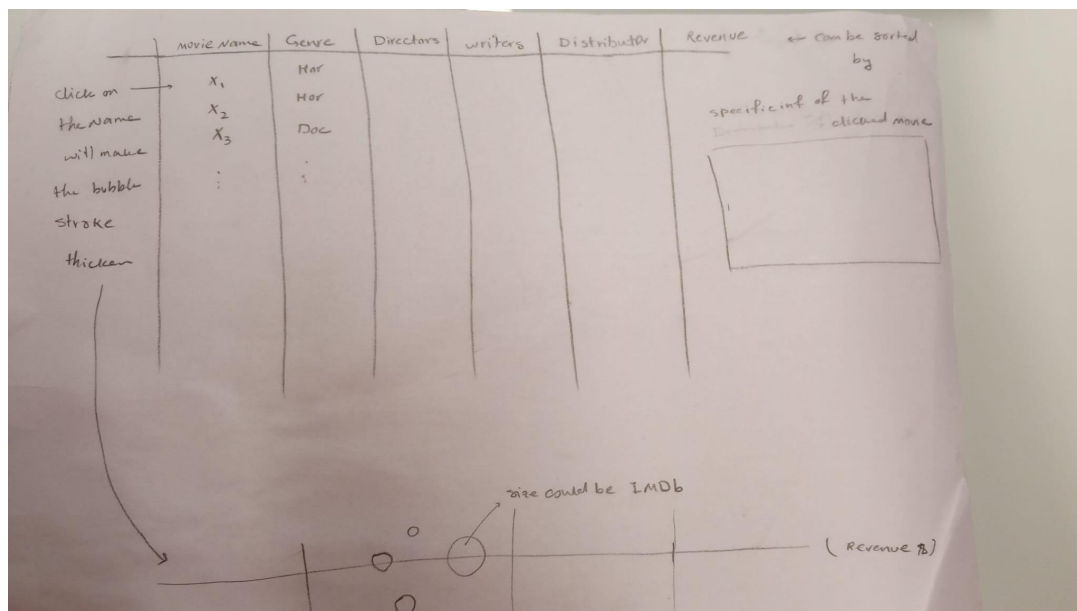
### **Dec 2: Final Project Due**

- Fix bugs
- Make minor fixes
- Complete process book
- Process book
- Code
- Project Website
- Project Screen Cast
- Peer Assessment

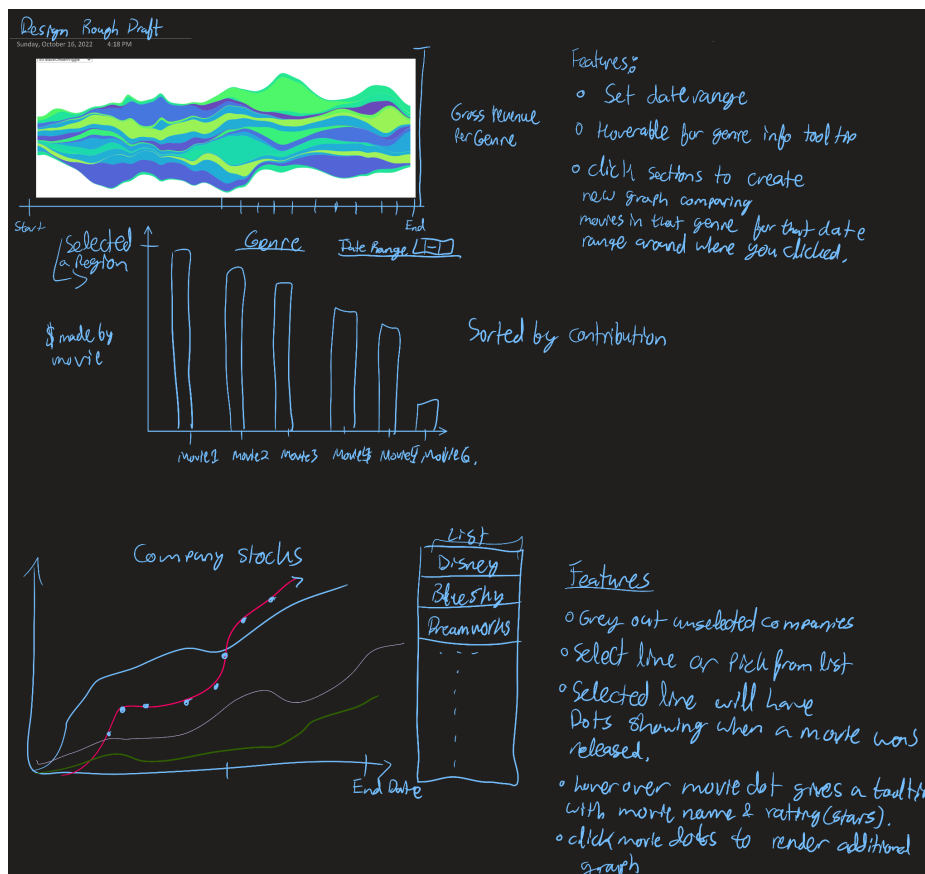
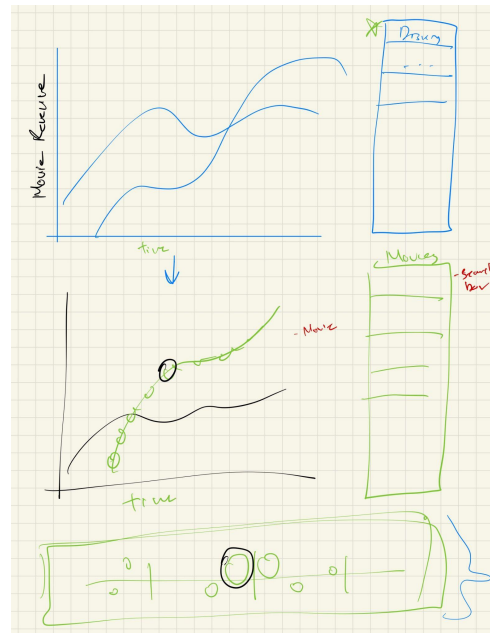
## Visualization Design Sketches:

Brainstorm: The following images show our initial brainstorming sketches.

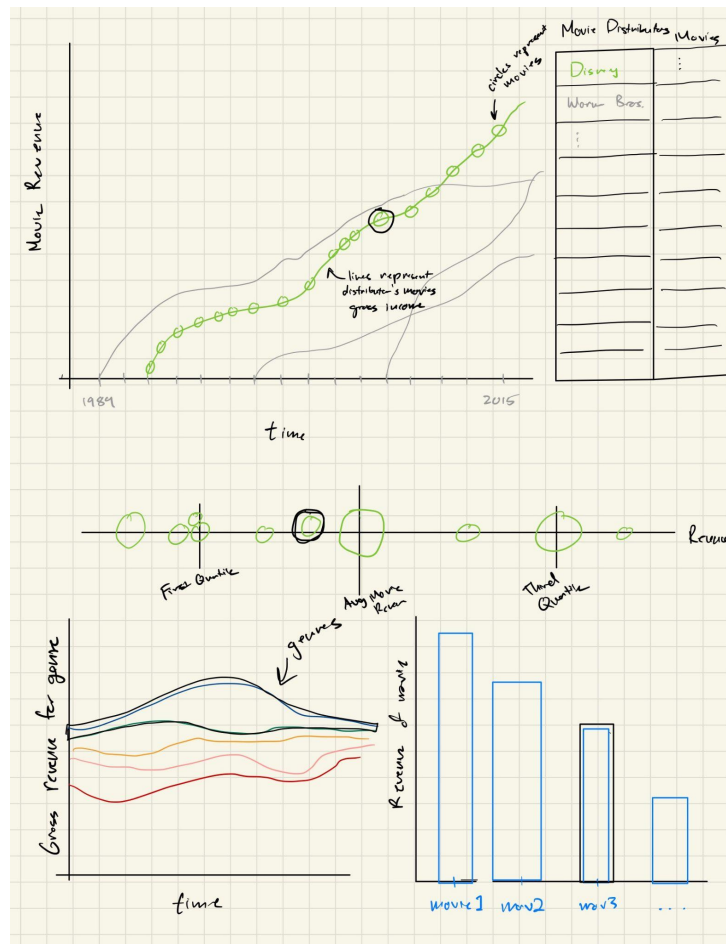




Initial Designs: We decided to use a line chart to show the cumulative company revenue with the gross income of the movies belonging to the company. After brainstorming, we scrapped scatter plots, since we have a lot of data points and felt it was not a good visual representation of our data. We want to display genres in some way and settled on a stream graph to do so. We also liked using the bubble chart to compare the IMDb score vs. movie revenue. Finally, we also will use a table to provide extra information for possible questions that a user may want to examine.

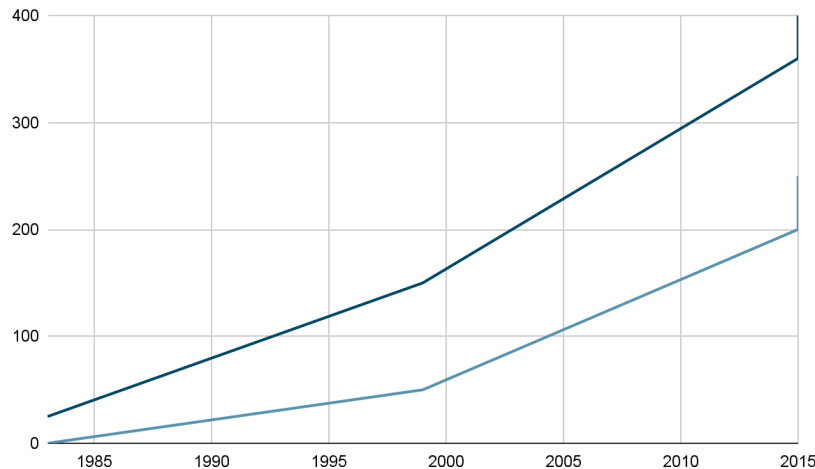


Realization Design: Our final design will feature three separate visualizations that all interact with each other.

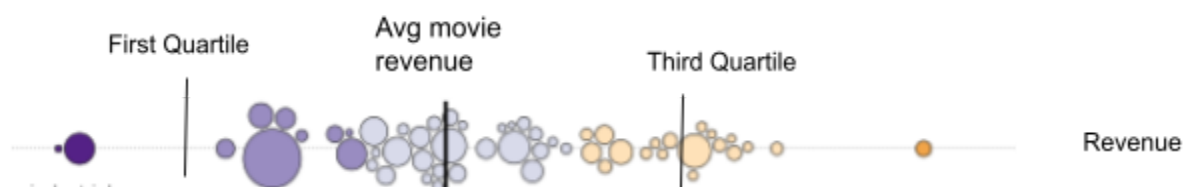
[illegible]



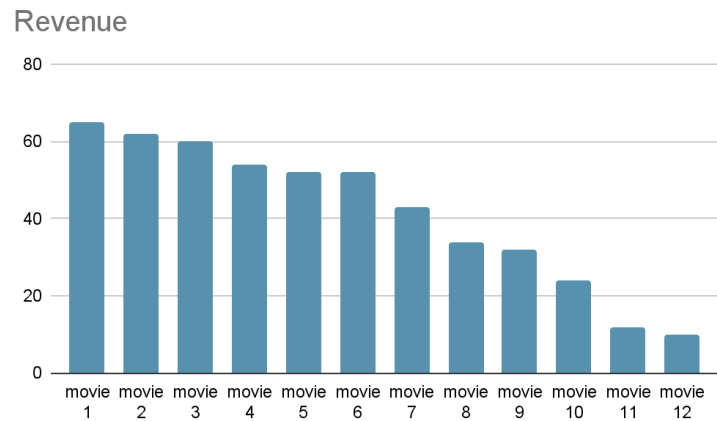
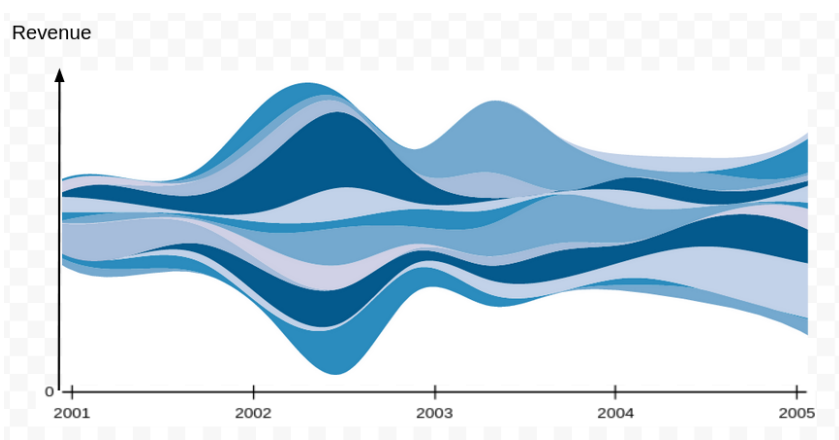
First, is a line chart that displays movie distributor data. The line is the cumulative movie income by each company. When a line gets selected, points get added on the line, each representing a movie. Additionally, a list of movies gets displayed alongside the list of distributors. Every other line gets grayed out.



As someone clicks on a single distributor, a bubble plot will display with all the movies of that distributor. It will show the average, first, and third Quartile revenue and the position of bubbles will show that movie's comparative contribution to the distributor revenue. Size of the bubbles show the movie IMDb score. This way we can investigate if the IMDb scores are correlated with the movie revenue and also we can see what movie(s) contribute to most revenues/ score



Then, a stream graph will be used to display the success of each genre over time. Each color will represent a genre over time with the y axis being revenue. The selection of a genre will result in a bar chart displaying 10 of the most successful movies of that genre sorted by revenue.



Finally, a table at the bottom of the page will display all extra data. It will get sorted on any event click on the page to display the most relevant data at the top. For example, when a genre is clicked, all movies with that genre will be displayed at the top of the table. When a singular movie is selected, the table will highlight that movie.

Movie Name	Distributor	Genre	Writer	IMBD rating	Revenue	...
Movie 1	Disney					
Movie 2	...					
Movie 3						
...						