

CS 6630: Visualizing Movies Metadata

Ram Kashyap S and Mohammed Musaddiq

Oct 28, 2017

1 Basic Info

Project Title: Visualizing Movies Metadata

Team member 1: Ram Kashyap S

Email: u1082810@utah.edu

uID: u1082810

Team member 2: Mohammed Musaddiq

Email: mohammed.musaddiq@utah.edu

uID: u1068996

Project repository:

<https://github.com/ramkashyap-s/dataviscourse-pr-movies-viz>

2 Background and Motivation

The reason we chose this project is the relevance and appeal it would have to a public audience considering the millions of movie fans across the world. Furthermore, being movie buffs ourselves, we are also motivated by personal interest in visualizing different aspects of movies and sharing the same fellow fans.

3 Project Objectives

The primary questions we are trying to answer with our visualization are:

- For a given actor/director, view or compare how the following parameters vary over time:
 - Rating of movies they have acted in/directed
 - Gross earnings of movies they have acted in/directed

- Budget of movies they have acted in/directed
- Number of movies they have acted in/directed per genre
- Proportion comparison of movies they have acted in/directed across genres

Benefits: The above visualization would allow users to get valuable insights into how an actor's/director's movies have fared over time based on various aspects they are interested about and wish to compare. It would also tell them about what kind of movies the actor/director is usually involved in.

- For a set of filters such as movie rating, year(s) and genre, provide a list of movies matching that criteria. Then, for a particular movie in this list, provide the ability to visualize other metadata such as Director, Actor, Language etc

Benefits: This visualization would help users in researching and finding all movies based on the genre(s) they like, how recent or old the movie is or how good/bad the movie's ratings are. Further, they can also view other potentially information about the movie they are interested in or are curious about.

4 Data

We have obtained the metadata for 5000+ movies spanning across 100 years in 66 countries from here:

https://github.com/sundeeblue/movie_rating_prediction/blob/master/movie_metadata.csv

The data contains 28 variables. There are 2399 unique director names and thousands of actors/actresses.

5 Data Processing

We expect to do a certain amount of data cleanup including:

- Check for empty/missing values
- Handle extraneous characters such as ^ in the movie titles column:
Avatar^
Pirates of the Caribbean: At World's End^
Spectre^

6 Visualization Design

Initial Design

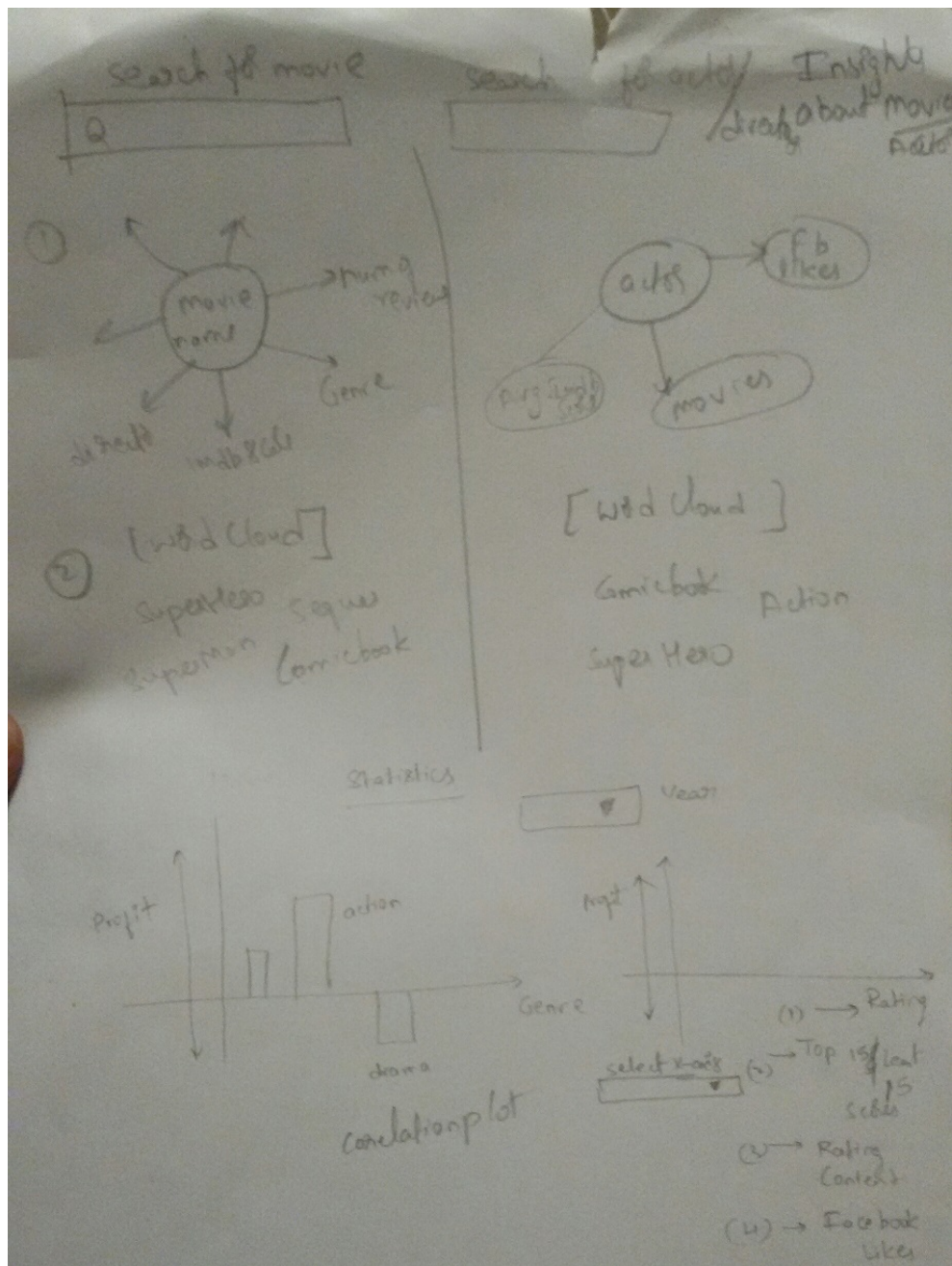


Figure 1: Initial Design

Implement a search filter that would allow the user to select a movie and an actor/director. Based on the selection, we will visualize that movie and actor/director as a graph with the movie and actor/director as central nodes and their associated metadata as child nodes attached to them.

We feel there might be a need to allow users to find/explore movies in other ways instead of restricting them to using a search filter that allows them to filter only by title.

We will also visualize the movie's plot keywords using a word cloud that would give the user some idea about what the movie is about.

Next, we will allow the user to select various movie attributes such as genre, rating using a drop-down list and visualize their correlation with movie profit using bar-chart.

Initial Design 2

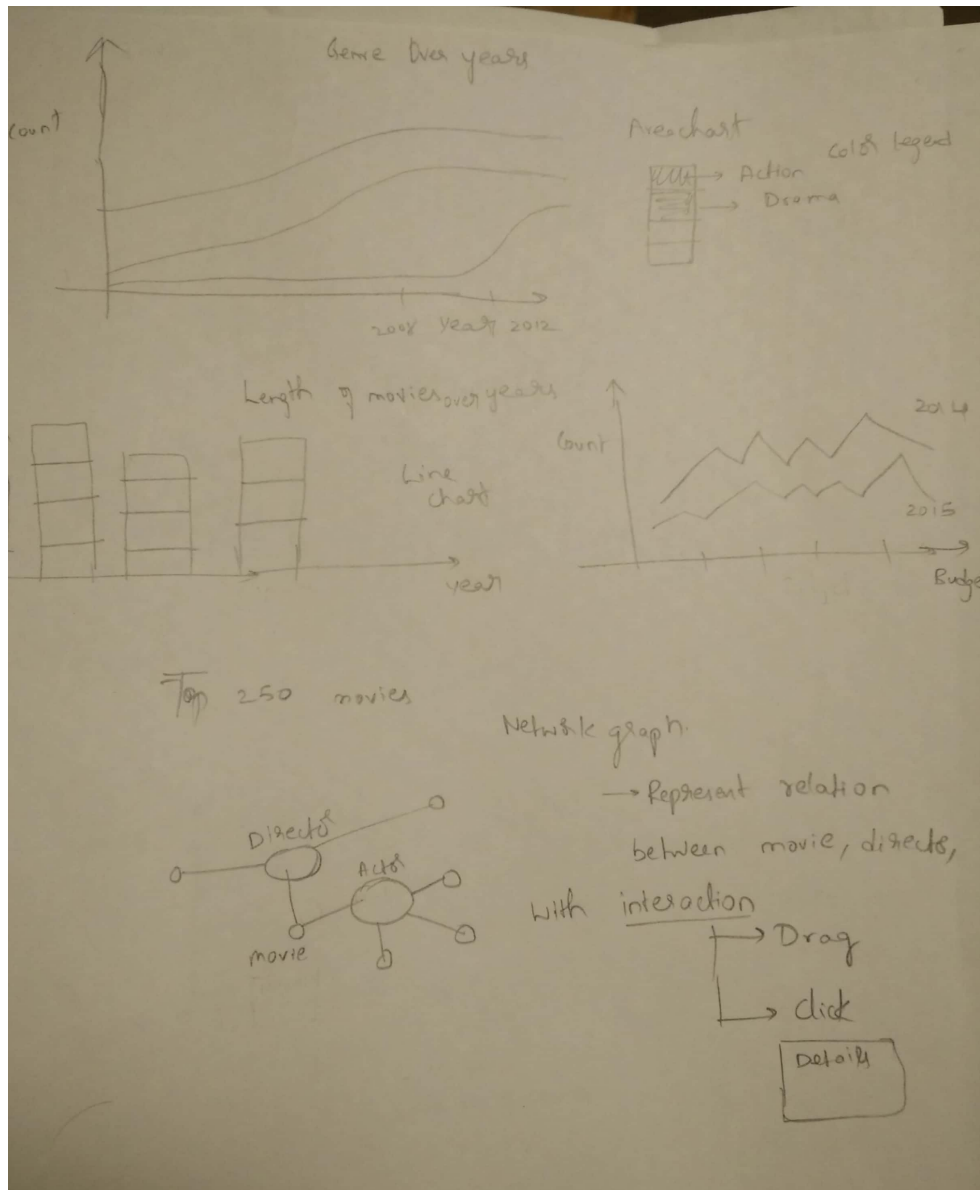


Figure 2: Initial Design 2

For top 250 movies implement a network graph which will represent relations between movie, director and actor. The size of the node for actor and director would be according to the number of connections(degree). The graph would be interactive with drag for visualizing the connections and click for visualizing the

details of node.

We feel that users might not get much information by looking at the connections in the graph.

Exploring data:

Implement area chart for genre trend over the years. Hue is used to encode different genres.

Implement a stacked bar chart for duration of the movies over the years.

Implement a line chart for budget over the years

Initial Design 3

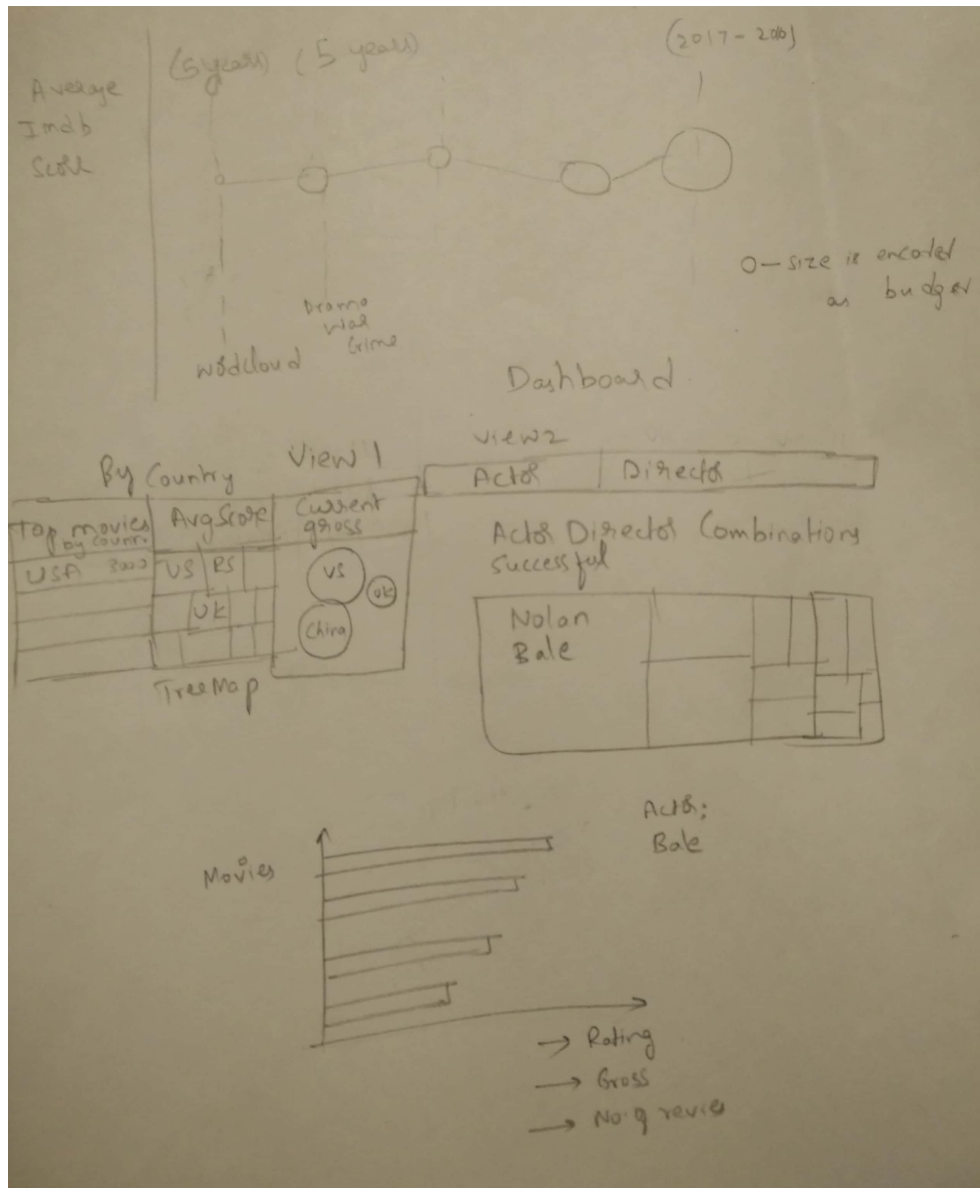


Figure 3: Initial Design 3

- Implement a dashboard with different views
- View by country: By each country display top twenty movies contributed
- Implement a tree map with average scores
- Implement a heat map of countries by selected year gross

View by actor and director:
Implement a tree map with successful actor and director combinations
Implement bar chart for top movies by rating, grossing, number of reviews,
etc. for an actor or a director

Finalized Design

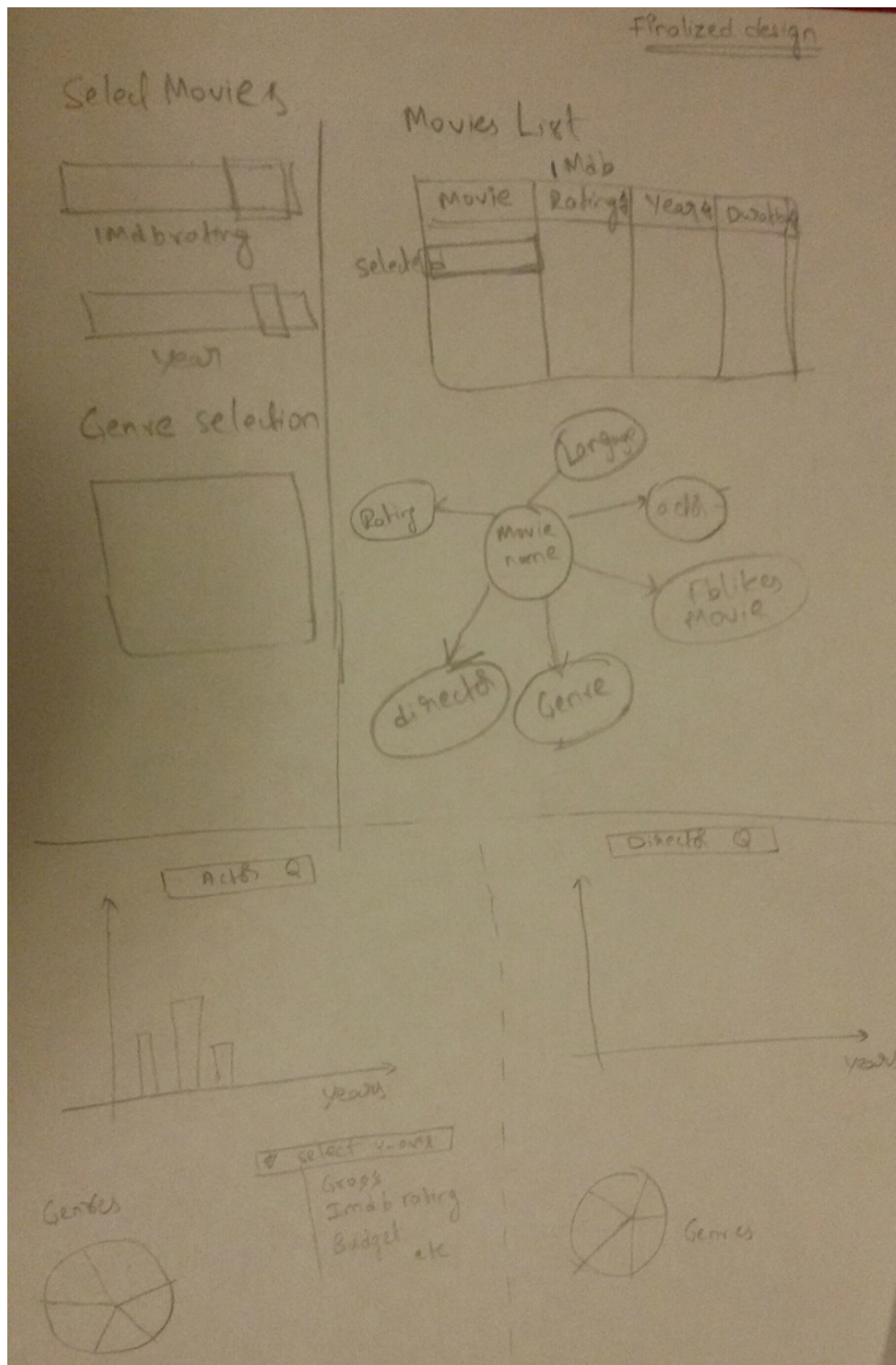


Figure 4: Finalized Design

We plan to implement a Movie Rating chart with a brush that would enable selection of a range of ratings that a user can specify as well as a brush-enabled Year chart that would allow the user to specify a range of years. We will also provide single/multiple genre selection using checkboxes.

We feel this is better than our initial design since earlier, the users did not have the ability to apply certain obvious filters to their search in case they did not know a movie name to begin with and just wanted to explore.

Then, we would visualize all movies matching the above criteria specified by the user using a dynamic table.

Finally, the user can select a movie from the above list he wants to know more about. We will visualize that movie as a graph with the movie as a central node and its associated metadata as child nodes attached to it.

Next, for the Actor/Director visualization, we plan to implement a search filter that would allow the user to search for and select an Actor/Director of interest. Based on the selection, we will provide a bar-chart based visualization of how different attributes of their movies have changed over the years. These attributes can be chosen using a drop-down list.

We will show the proportion of movies across various genres for the selected Actor/Director using a Pie Chart (and maybe encode the number of movies as a label inside the associated slice/sector or as a tooltip, if needed). Although a pie chart is usually not recommended, we feel it is well suited for visualizing this particular information in a simple and effective manner.

7 Must-Have Features

We feel that all the features mentioned in the **Finalized Design** should be considered as a Must-Have.

8 Optional Features

Some nice to have features would be:

- word cloud visualizing a movie's plot keywords
- display a selected movie's poster or actor's/director's photo

9 Project Schedule

Module-wise responsibilities:

- Ram - Implement the movie filter, selection and visualization flow
- Mohammed - Implement the actor/director selection and associated analysis flow

Week 1: Plan the overall code structure and write skeleton html, javascript classes/functions required for implementing the various modules.

Week 2 - 3: Implement all the required modules outlined in the **Finalized Design**

Week 4: Code cleanup, testing, bug fixes (if any). Also try implementing enhancements or optional features if time permits.

Week 5: Finish any remaining process book updates. Project hosting and screen-cast.