

The background of the slide features a faded, artistic image of film strips. The strips are arranged in a circular, overlapping pattern, creating a sense of motion and depth. The film strips are a light gray color, contrasting with the white background.

cs171 - Final Project

Ceruti, Mychal

PROCESS BOOK

2014.04.08

The background of the slide features a blurred, overlapping image of film strips, suggesting a cinematic or archival theme.

CS171 | FINAL PROJECT

TABLE OF CONTENTS

The background of the slide features a close-up, slightly blurred image of several film reels. The reels are arranged in a circular pattern, with their edges and the perforations of the film strips visible. The lighting is soft, creating a cinematic and artistic feel.

CS171 | FINAL PROJECT

HOLLYWOOD ACTOR'S BANKABILITY OVER TIME (1993-2013)

This visualization provides you with an interactive and engaging visualization showing the rise, fall, and overall performance of the most popular Hollywood actors and actresses over the last 20 years.

The background of the slide features a blurred, artistic image of several overlapping film strips. The strips are light gray with visible sprocket holes, creating a sense of motion and cinematic history. They are arranged in a way that they seem to be part of a larger, continuous reel.

PROPOSAL | INSPIRATION

RELATED WORK

The Hot List

Most everyone has heard of the A,B,C, and D-list. The terms were created by veteran journalist James Ulmer to gauge the bankability of actors and directors...

The background of the slide features a close-up, slightly blurred image of several film reels. The reels are arranged in a circular pattern, with their edges and the perforations of the film strips visible. The lighting is soft, creating a vintage and cinematic atmosphere.

PROPOSAL | QUESTIONS

PROPOSAL | PROBLEMS

EARLY ISSUES

Difficult to Quantify

Although there exists a wealth of data available online on movies, including cast lists, rating, budgets et cetera, finding consistent, historical aggregated data that quantify an actor's talent, popularity, and other attributes that are essential to determining their bankability (defined by The Hot List's page as: “*the degree to which an actor's name alone can trigger full financing for a movie.*”). Finding this data has proven to be a hurdle that may be impossible to overcome with the time given financial and time constraints. As a result, I will still endeavor to create a visual tool that can effectively compare celebrities' careers over time.

Data is Everywhere (and they want you to pay for it)

To determine an actor's ability to “put butts in seats”, there are several factors that have to be taken into consideration. The Ulmer scale, for example takes into account talent, ...

Change of Direction

For Design Studio 3 we were tasked with providing and receiving feedback from other groups as to the design and scope of our project. During my interview with Yash, he pointed out that while interesting, my goals may have been a bit ambitious for what would be expected from a one-man-team. After a short brainstorming session, we had come up with a solution that, while moving away from my original vision, would still achieve what I had set out to accomplish, namely a fun, engaging tool that would show the development of an actor or actors over a span of time in their careers. The solution allowed me to use much of the code I had written for my original designs, and I was introduced to another source of movie data that was robust enough and accessible enough to grab all the data I needed for my visualization.

I will be leaving in my previous methods and the type of data that I had grabbed so far for the purposes of showing my process, but as of 2014.04.10, I've decided to make a fairly major change as far as the scope of my project. Despite the new direction however, the visualization will still accomplish what was the primary objective of my project

PROPOSAL | CHANGE OF DIRECTION

Revised Format




Revised Sketch statistics for actor

PROCESS | DATA

DATA COLLECTION

After I had a rough idea for what I wanted to accomplish, I began my search for sites where I could gather the data that I needed. IMDB and RottenTomatoes.com both provided APIs that allowed access to much of their data, but scraping data some of the other websites proved to be a little more difficult.

Lee's Movie Info: I needed to find an easy way to gather lists of the top movies from last 20 years. My first choice was Wikipedia or IMDB, but as I was still wrapping my head around the use of the IMDB API and writing code that could gather that information from Wikipedia would have proven too time consuming I settled on manually wrangling the data from Lee's Movie Info. Lee's Movie Info has its top 50 list of movies sorted in tables that were easy to copy and paste into a spreadsheet which I then converted to a CSV file. Though the site had a lot of other valuable information, I couldn't find an API that granted me easy access to it so I decided to look elsewhere.



The screenshot shows the 'LEE'S MOVIE INFO' website. At the top, there are navigation tabs: BOX OFFICE, PREDICTIONS, REVIEWS, UPCOMING, and MOVIE TALK. Below these, a year selector shows '1965 | 1970 | 1975 | 1980 | 1985 | 1990 | 1995 | 2000 | 2005 | 2010'. A sub-header reads '<< 2008 Top Movies Released In 2009 2010 >>'. The main table lists the top 50 movies with columns for Rank, Title, Total, Budget, and Studio. On the left side of the page, there is a search box labeled 'Movie Title' with a 'Find' button, and a sidebar menu with categories like 'BOX OFFICE', 'Daily Box Office', 'Weekend Box Office', 'Weekly / Monthly', 'Decade', 'All-Time Box Office', and 'Breakdown'.

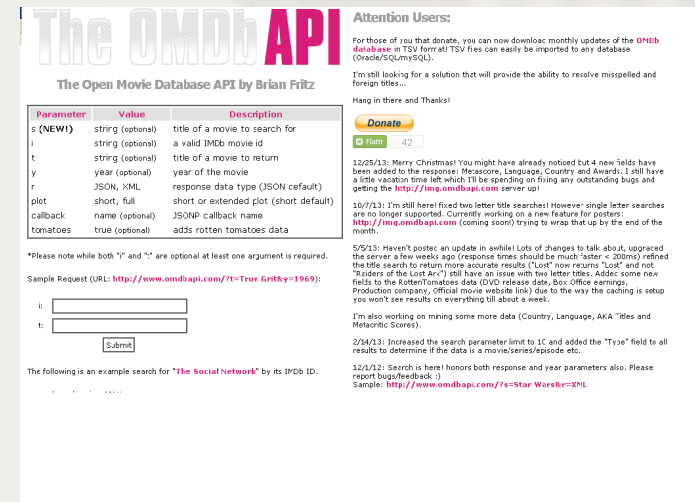
Rank	Title	Total*	Budget*	Studio
1	Avatar	\$759.563	\$190.0	20th Century Fox
2	Transformers 2	\$402.077	\$200.0	Paramount
3	Half-Blood Prince	\$301.957	\$250.0	Warner Brothers
4	Twilight: New Moon	\$293.815	\$51.0	Summit
5	Up	\$292.913	\$175.0	Buena Vista
6	The Hangover	\$277.313	\$35.0	Warner Brothers
7	Star Trek	\$257.704	\$150.0	Paramount
8	The Blind Side	\$255.945	\$29.0	Warner Brothers
9	Alvin and Chipmunks 2	\$219.607	\$75.0	20th Century Fox
10	Sherlock Holmes	\$208.979	\$90.0	Warner Brothers
11	Monsters vs. Aliens	\$198.33	\$175.0	DreamWorks
12	Ice Age 3	\$196.44	\$90.0	20th Century Fox
13	Wolverine	\$179.86	\$150.0	20th Century Fox
14	Night at the Museum 2	\$177.23	\$150.0	20th Century Fox
15	2012	\$165.55	\$200.0	Columbia
16	The Proposal	\$163.90	\$40.0	Disney
17	Fast and Furious	\$154.88	\$85.0	Universal
18	G.I. Joe	\$150.17	\$170.0	Paramount
19	Paul Blart: Mall Cop	\$146.34	\$25.0	Columbia
20	Taken	\$145.00	\$25.0	20th Century Fox
21	A Christmas Carol	\$137.85	\$200.0	Disney
22	Angels & Demons	\$133.38	\$150.0	Columbia
23	Terminator Salvation	\$125.29	\$200.0	Warner Brothers
24	Chance of Meatballs	\$124.87	\$100.0	Columbia
25	Inglourious Basterds	\$120.52	\$70.0	Weinstein Co.
26	G-Force	\$119.25	\$150.0	Buena Vista
27	District 9	\$115.65	\$30.0	Columbia
28	It's Complicated	\$112.70	\$85.0	Universal

Lee's Movie Info: Top Movies of 2009

PROCESS | DATA

OMDB API: As IMDB is one of the largest, most comprehensive, consistent, and free databases of film data, it was an obvious choice. While it does have a publicly accessible API, it doesn't have any official support or documentation. Thankfully, there are many websites that provide their own methods for accessing IMDB to collect data. I used OMDB API(<http://www.omdbapi.com/>).

Java Movie Database: IMDB does not allow screen scraping, and all of the public APIs I found that accessed IMDB's didn't provide methods for scraping the salary data, an essential component of my project. IMDB provides plain-text versions of most of their data, but the files are large and difficult to parse through. Because of this I searched for a tool to make the job easier and came across the Java Movie Database (JMDB). JMDB loaded all of the plaintext data into a database on my local server and provided a GUI that allowed me to search and export all of the information I needed. Though this was ultimately more tedious than gathering information programmatically (which would have been possible, but I felt I didn't have sufficient time to learn and write code that searched through the databases), it got the job done.



The OMDB API

Attention Users:

For those of you that donate, you can now download monthly updates of the OMDB database in TSV format! TSV files can easily be imported to any database (Oracle/SQLMySQL).

I'm still looking for a solution that will provide the ability to resolve misspelled and foreign titles...

Hang in there and Thanks!

Donate

Funds 42

12/25/13: Merry Christmas! You might have already noticed but 4 new fields have been added to the response: Metascore, Language, Country and Awards. I still have a little vacation time left which I'll be spending on fixing any outstanding bugs and getting the <http://img.omdbapi.com> server up!

10/7/13: I'm still here! Fixed two letter title searches! However, single letter searches are no longer supported. Currently working on a new feature for posters: <http://img.omdbapi.com> (coming soon) trying to wrap that up by the end of the month.

5/5/13: Haven't posted an update in awhile! Lots of changes to talk about, upgraded the server a few weeks ago (response times should be much faster < 200ms) refined the title search to return more accurate results ("Lost" now returns "Lost" and not " Raiders of the Lost Ark") still have an issue with two letter titles. Added some new fields to the RottenTomatoes data (DVD release date, Box Office earnings, Production company, Official movie website link) due to the way the caching is setup you won't see results on everything till about a week.

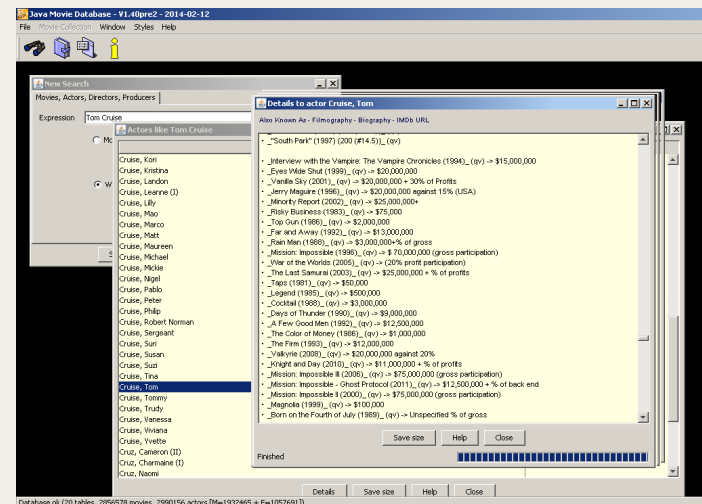
I'm also working on mining some more data (Country, Language, AKA titles and Metascore Scores).

2/24/13: Increased the search parameter limit to 1C and added the "Type" field to all results to determine if the data is a movie/series/episode etc.

12/2/12: Search is now honors both response and year parameters also. Please report bugs/feedback :)

Sample: <http://www.omdbapi.com/?t=Star Wars&r=3XIL>

OMDB API Homepage



Java Movie Database - V1.40pre2 - 2014-02-12

File Movie Controls Window Styles Help

Movie Search: Tom Cruise

Expression: Tom Cruise

Movie: Tom Cruise

Director: Tom Cruise

Producer: Tom Cruise

Details to actor Cruise, Tom

Also known As - Filmography - Biography - IMDb URL

- "South Park" (1997) (TV) (#14.5) (TV)
- Interview with the Vampire: The Vampire Chronicles (1994) (TV) -> \$15,000,000
- Eyes Wide Shut (1999) (TV) -> \$20,000,000
- Vanilla Sky (2001) (TV) -> \$20,000,000 + 50% of Profits
- Jerry Maguire (1996) (TV) -> \$20,000,000 against 15% (USA)
- Money Report (2002) (TV) -> \$25,000,000+
- Dirty Business (1987) (TV) -> \$75,000
- Top Gun (1986) (TV) -> \$2,000,000
- Far and Away (1992) (TV) -> \$13,000,000
- Rain Men (1989) (TV) -> \$3,000,000-% of gross
- Mission: Impossible (1996) (TV) -> \$ 70,000,000 (gross participation)
- Star of the Vindicta (2005) (TV) -> 20% profit participation
- The Last Samurai (2003) (TV) -> \$25,000,000 + % of profits
- Ties (1981) (TV) -> \$50,000
- Legend (1985) (TV) -> \$500,000
- Cocktail (1988) (TV) -> \$3,000,000
- Days of Thunder (1986) (TV) -> \$3,000,000
- A Few Good Men (1992) (TV) -> \$12,500,000
- The Color of Money (1986) (TV) -> \$1,000,000
- The Firm (1993) (TV) -> \$12,000,000
- Velocity (2003) (TV) -> \$20,000,000 against 20%
- Knight and Day (2010) (TV) -> \$11,000,000 + % of profits
- Mission: Impossible II (2003) (TV) -> \$12,500,000 + % of back end
- Mission: Impossible - Ghost Protocol (2011) (TV) -> \$12,500,000 + % of back end
- Magnolia (1999) (TV) -> \$10,000
- Born on the Fourth of July (1989) (TV) -> Unspectored % of gross

Save size Help Close

Database ok (20 tables, 2096578 movies, 2990156 actors [N=1932465 + F=1057691])

Java Movie Database

PROCESS | DATA

CLEANUP AND STRUCTURE

The revised visualization required much more compact data, and was more easily accessible and can allow for a more dynamic graph. The original data structure used 2 sets of data, one for movies and one for actors. ...

The most recent structure has...

OLD STRUCTURE

Dataset 1

```
movie:{
```

Dataset 2

NEW STRUCTURE

Dataset 1

```
[movie:{grossDomestic:112, grossGlobal:238.2, sum:350.2,
actors:[array](list of actors),releaseDate:{dateobject},
poster:"http://www.bar.com/poster.jpg", movie2:{gross...}]
```

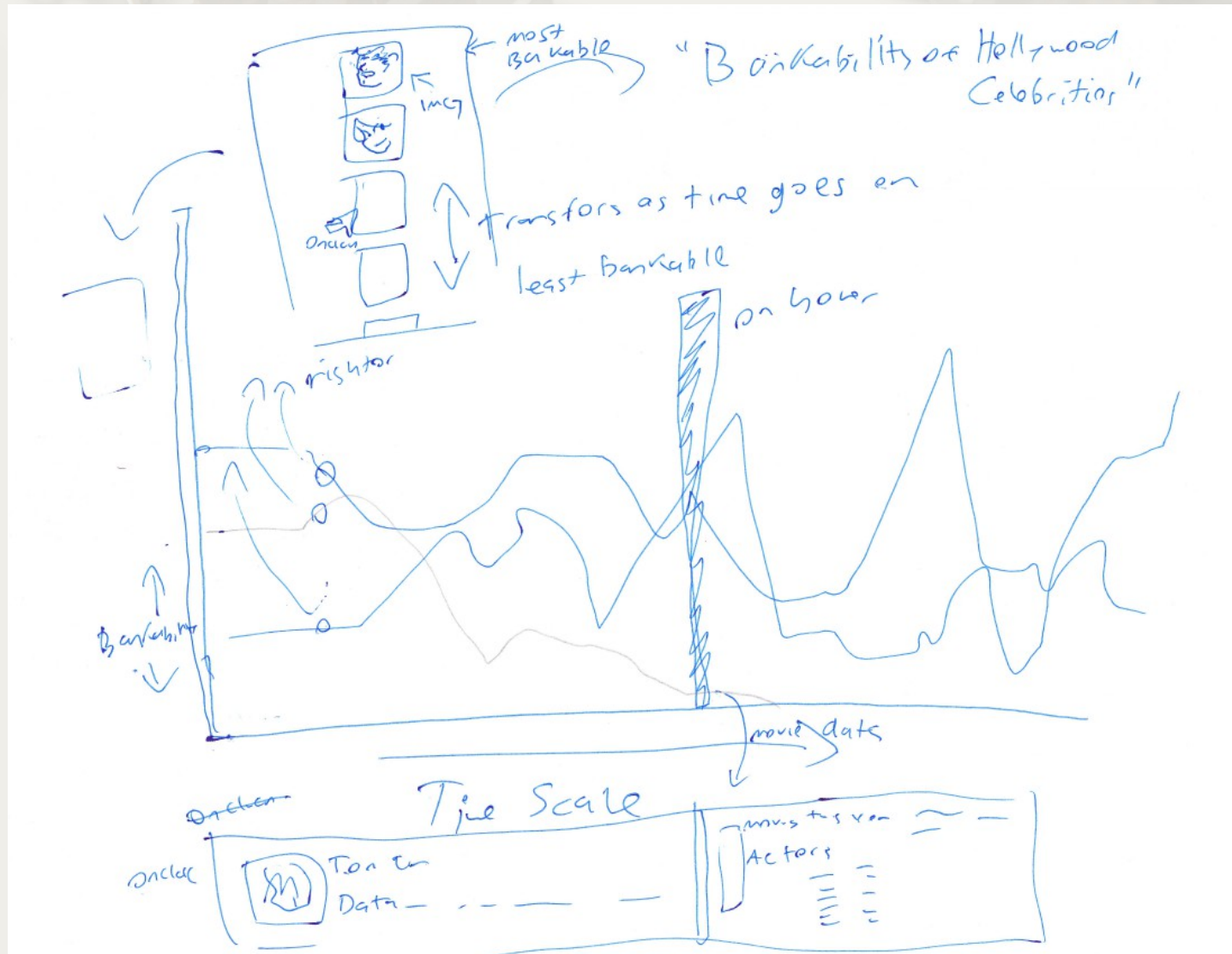
Dataset 2

```
[actor:{careerGross:409.3, image:"
http://www.foo.com/actor.jpg", movies:[array](list of
movies)}, actor2:{careerGross... }]
```

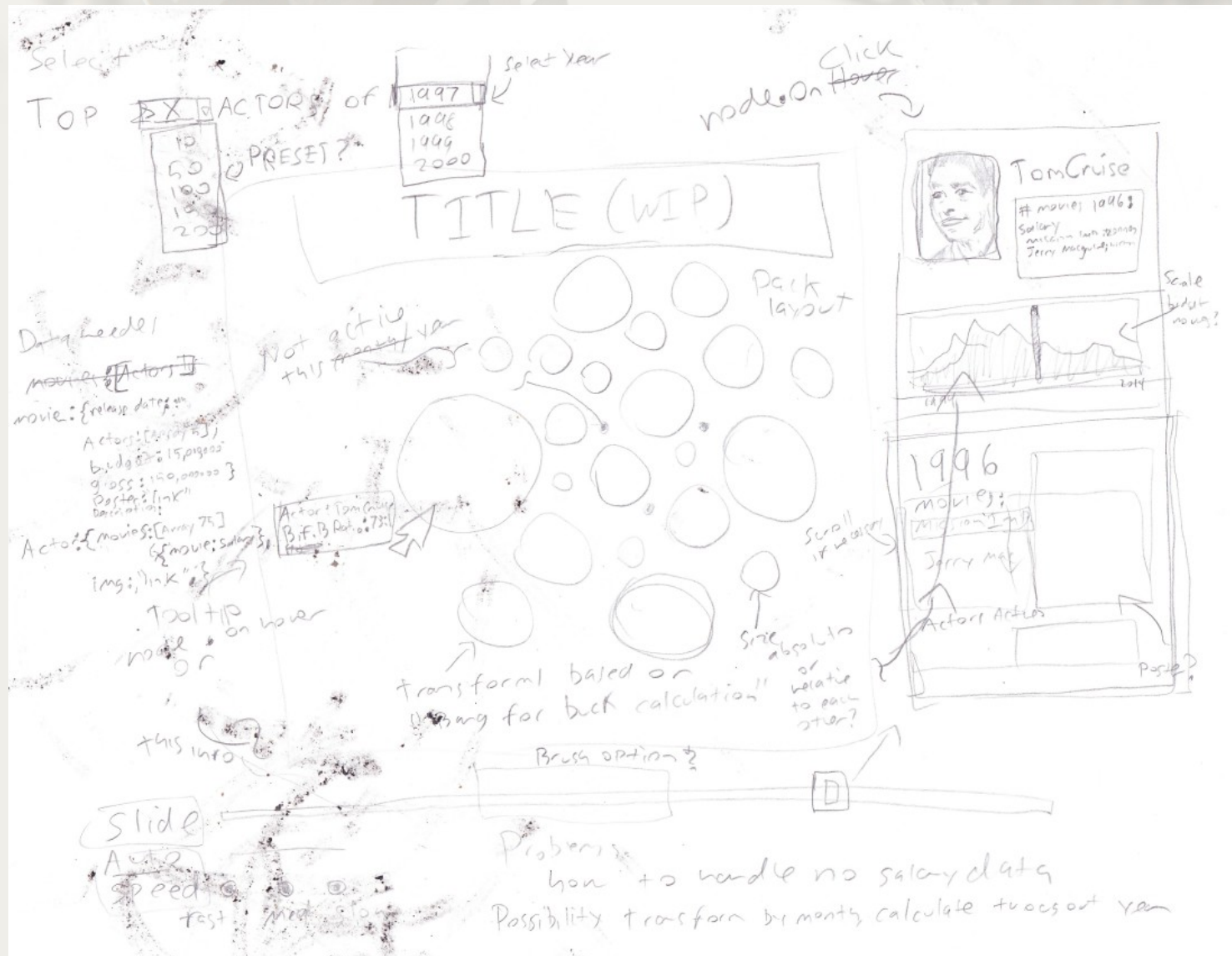
The background of the slide features a close-up, slightly out-of-focus image of several film reels. The reels are arranged in a circular pattern, with their edges and the perforations of the film strips visible. The lighting is soft, creating a professional and artistic feel.

PROCESS | ANALYSIS

VISUALIZATION | DESIGN



VISUALIZATION | DESIGN



The background of the slide features a close-up, shallow depth-of-field shot of several interlocking film reels. The reels are metallic and show the characteristic sprocket holes. The lighting is soft, creating a vintage, cinematic atmosphere. The text is overlaid on the upper portion of the image.

VISUALIZATION | DESIGN

The background of the slide features a close-up, shallow depth-of-field photograph of several film reels. The reels are arranged in a circular pattern, with the focus on the foreground reels, making the background reels appear blurred. The lighting is soft, highlighting the metallic texture of the reels and the perforations along the edges of the film strips.

VISUALIZATION | FEATURES

The background of the slide features a close-up, slightly out-of-focus image of several film reels. The reels are arranged in a circular pattern, with their edges and the perforations of the film strips visible. The lighting is soft, creating a vintage and artistic feel.

VISUALIZATION | ANALYSIS

The background of the slide features a close-up, slightly out-of-focus image of several overlapping film strips. The strips are dark with visible sprocket holes and some text, creating a textured, cinematic feel.

CONCLUSIONS

- A few takeaways I've

The background of the slide features a close-up, slightly out-of-focus image of several overlapping film strips. The strips are dark with visible sprocket holes and some text, creating a textured, cinematic feel. The word "BIBLIOGRAPHY" is centered over this background in a large, black, serif font.

BIBLIOGRAPHY