MOVIES AND ACADEMY AWARDS - FINAL PROJECT REPORT

Team Members: Nia Vivek, Priyanka Patki and Astika Gupta

1. Project goals and tasks the interface containing the visualization is targeted towards

To create a fun and interactive tool that allows movie enthusiasts to play around with the different visualizations to explore and find out more about their favorite movies, movie stars and directors present in the IMDB and academy awards dataset.

Some ideas that movie lovers could explore or discover are if there is any correlation between the movie IMDB score and budget of their favorite movies or actors and number of academy awards that those movies have been nominated/won, movie genres and award winning actors and director pairing with award winning actors.

We hope to allow users to explore and find out if there is a dominant trend or pattern that emerges through our visualizations. Our visualization is focussed on users who watch movies from once a week to once every two weeks or more and who consider the ratings, critics reviews, actors and director of the movie before watching it.

Finding the right direction - Conducting a survey:

We conducted a survey asking few direct questions to gauge better what interests people more when it comes to movies and academy awards, what is it that they would like to know more about, what are their movie watching habits etc. Taking into consideration all of these factors and the responses that we got, we could better narrow down our focus on the aspects that we wanted to show through our visualizations.

2. Discussion of related work

We started off by going through a couple visualizations related to our topic- movies to get more inspiration and a starting ground to get ideas and directions.

Two major ones that we consulted are as follows-

a. https://public.tableau.com/en-us/s/gallery/imdb-movies-visualized - This visualization can provide us with inspiration for interaction. It has the capability of searching free form text, filtering, selecting different genres and countries and then exploring the results. It is quite interactive and gives a lot of freedom to the user to choose what combination he/she wants to explore. Though it provides interaction, it uses a single type of graph

with a bunch of information in the tooltip and we feel a better design would show the information more interestingly.

b. http://gigapan.com/gigapans/4306 - Though this visualization has explored interesting aspects like all of the actors who ever acted together but the design of this visualization is quite poor and not user-friendly. The design is quite messy and has all the titles embedded which does not give the users any useful information. Some of the color choices are bad like grey text on white background is completely lost. There is no interactive feature that allows more exploration.

One of the struggles that we faced was given our large dataset, our first and foremost step was to somehow select the columns that we wanted to focus on and this meant deciding on what it is that we really want to show through our visualizations in a way that not only is it interactive but also covers as many aspects related to movies and awards as possible. We then decided to focus on one visualization at a time and get inspiration for that individually. Following is the list of related works that we consulted for each of the 3 visualizations.

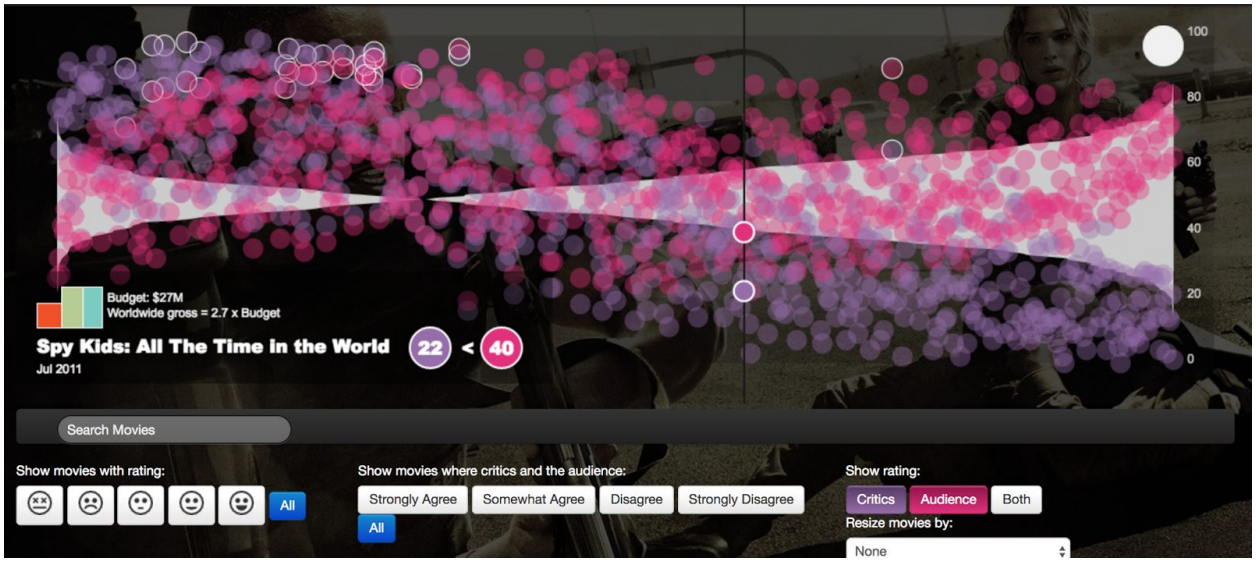
Inspirations for Visualization 1:

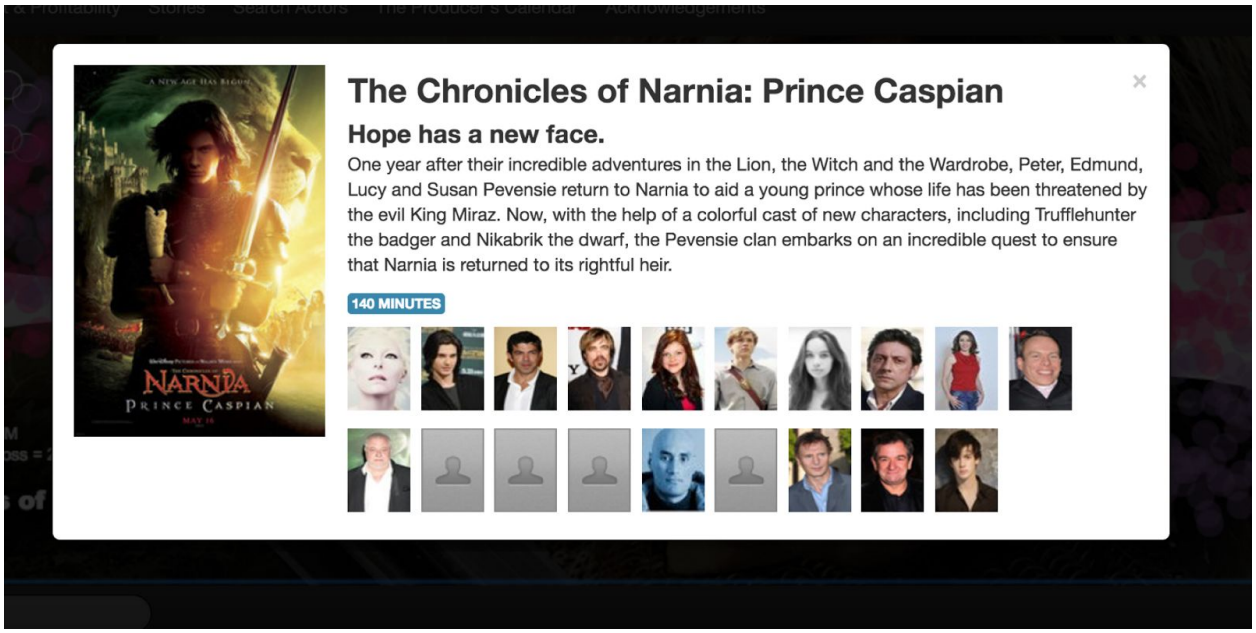
Right at the beginning of the visualization, we wanted to give the user a chance to play around with some numbers and give them a chance to explore various parameters that are integral to movies.

Looking at our data, given our area of interest i.e movies and awards and from the response we received from the survey we conducted and after conducting some exploratory analysis in Tableau, we decided we wanted to incorporate at least some measures like budget, movies, awards, genres etc. in some form in our visualization. Keeping in mind the volume of data we had, we wanted to keep a familiar visualization like a scatter plot showing multiple forms of data in the beginning. For our initial design we thought of incorporating length or position, size and color as some of the ways of representing data. Thus we began to explore other visualizations that visualized these numbers for various movies.

Related work-1:

Confluence:





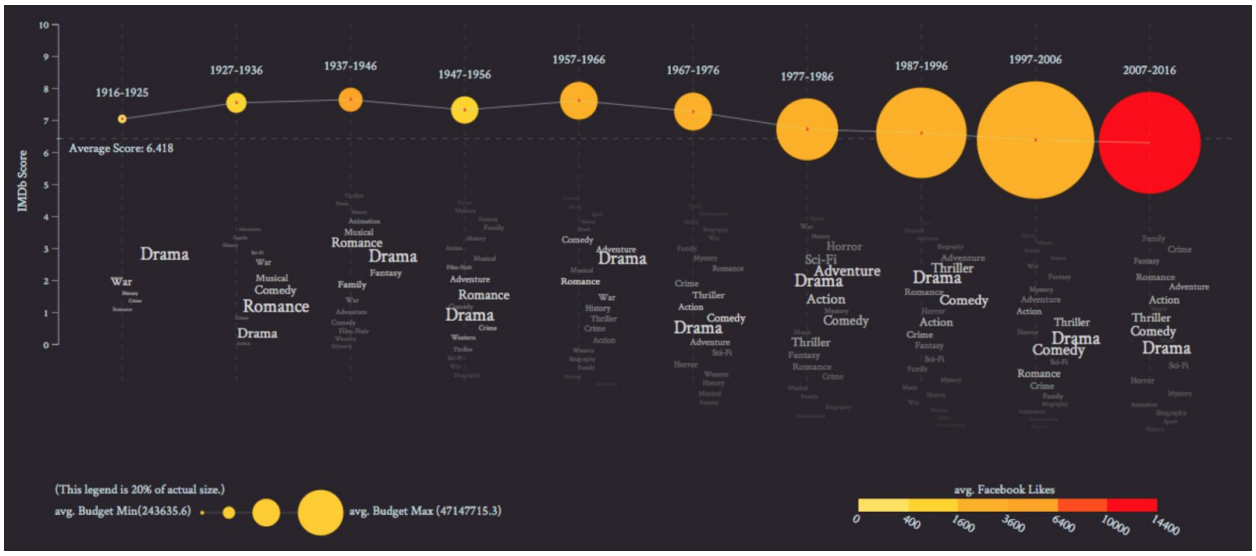
http://iibh.apphb.com/

This visualization was one of the first visualizations we encountered when we were looking for some inspiring info visualizations that were done in the movies domain. This visualization basically helped to visualize how difference in opinion among audience

and critics can be shown along with other parameters like budget. The different colors i.e pink and purple signify audience opinion and critics opinion respectively and the size of each bubble changes based on the parameter selected like –budget, gross revenue etc. They also had tabs where each tab would allow the user to filter and search data based on different parameters. There were also some other filter criteria that were standard and were common among all the visualizations. Also, upon clicking a bubble the movie details like posters, actor details, plot summary etc for that movie popped up in another small window. All in all, this visualization did a great job in portraying how difference in public and critic opinion changed and how different parameters like budget, story shaped the visualization. However, there were still many shortcomings of this visualization. Although, this visualization was great to slice and dice data on many levels, the lack of clear legends for X and Y axis made it initially difficult to figure out what exactly was being shown without reading the description given below it. There was also a white area highlighted in the chart that demonstrated how for different movies the difference in opinion between critic score and audience opinion changed. This at first was not apparent and it took some time and exploration on our part to realize what the designer was trying to achieve with this visualization. Additionally, as there was a lot of data that was being visualized, without the option of zooming or concentrating on a given the area, it was difficult to find any distinct trend when additional parameters like budget, gross revenue etc. were being visualized on top of the basic graph of difference in opinion graph.

Thus, being one of the first visualizations that we encountered that talked about movies in terms of critics and audience popularity along with many other measures bolstered our idea and gave us more confidence to explore the option of visualizing movies in terms of critics score and audience score showed through meta critic score, IMDB score and budget respectively in our visualization. Initially we thought of incorporating tabs, like the this visualization, that gave the user the ability to navigate different filters – like movies tab would allow the user to search based on movies, actors tab could allow search based on actors, however after getting some feedback on our initial designs through usability testing we realized that keeping the filters on separate tabs does not give the user a chance to filter data based on these different parameters together. During our initial design talk we did want to include movie posters or actor photos and other details about the movie in some form and looking at this visualization strengthened our idea to incorporate movie posters and movie details in a tool tip, or which we later changed to be shown on the right side of the visualization.

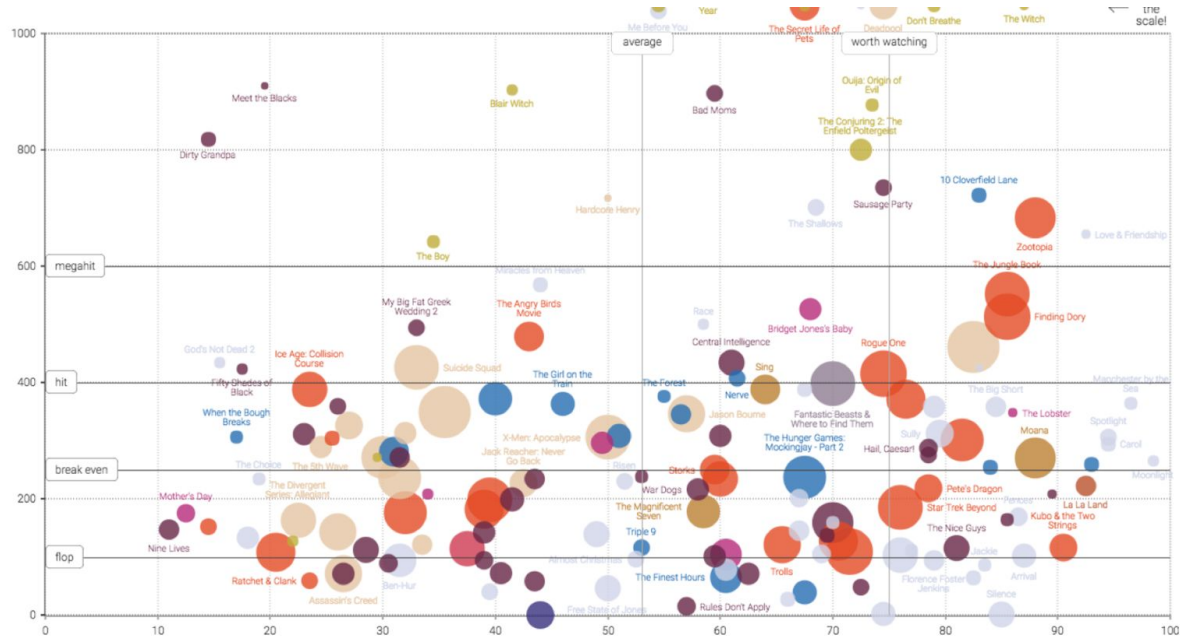
Related work 2:



https://medium.com/towards-data-science/the-century-of-movies-12a4b5a4c004

This was another visualization done for movies that involved combining multiple forms of information like IMDB score, genres, year and Facebook likes. This was another great way of representing lot of data nicely through a combination of line chart and a bubble chart. The size of the font was also nicely used to show the dominant genres in those years and the clever and limited use of color gradient were nicely incorporated to depict facebook likes with respect to time. Thus by plotting grouped data, it was certainly easier to make more sense of the data, specially the kind of data that was represented through color and size. Thus, this visualization forced us to think that while we found that it is important to provide users with a lot of data to explore, it is also important to give user a chance to see limited data. It also showed us that there were some interesting trends to be explored with movies and genres and the budget they were made in.

Related work -3:



http://www.informationisbeautiful.net/visualizations/the-hollywood-insider/

Looking at another visualization that covered budget and critic ratings in movies, reinforces our idea of incorporating multiple aspects in the same visualization. This visualization does a great job of showing budget recovery, gross revenue, average critics %, genre and budget. This visualization basically is an exploratory visualization that allows user to explore data for about 1200 movies in the last ten years. The use of color not only allows the viewer to see information genre but also makes the design appealing. Seeing how size and color is used and that it was not too difficult to read and incorporate a lot of data in many forms on the same plot, alleviated some of our concerns that incorporating a lot of information by representing the data in multiple forms like size, color etc. can be very difficult to read. Thus, it further strengthened our idea to focus on representing data through size and color on the same scatter plot. Through this chart we also realized that being able to filter down information would allow user to see the data and what the numbers, size, shape and color meant more clearly. Thus, it gave us the idea of incorporating filters or drop down menus to give the users the ability to explore the data more.

Inspirations for Visualization 2:

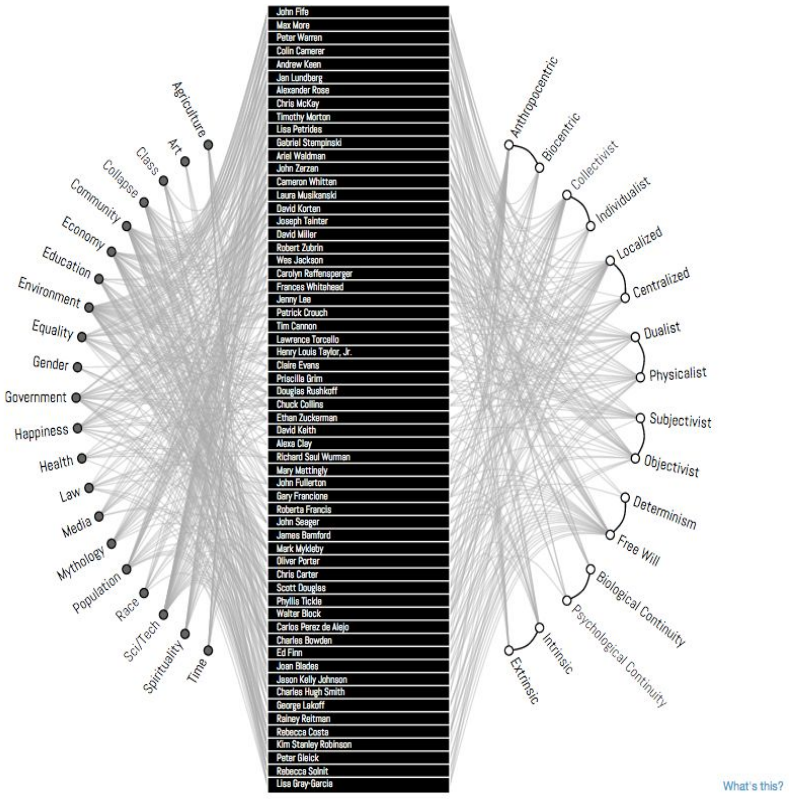
For the second visualization (The concept map) that maps the relationship between movie genres and the award winning actors, we studied a number of visualizations online to get inspiration. Our end goal was to be able to show clearly the individual mapping between the actors who have won awards and the genres of the corresponding movies for which they have won the awards. Also, we wanted to show this relationship both ways - the movie genres related to a particular actor and the actors related to a particular genre.

i. Related work 1: *Find The Conversation Concept Map*

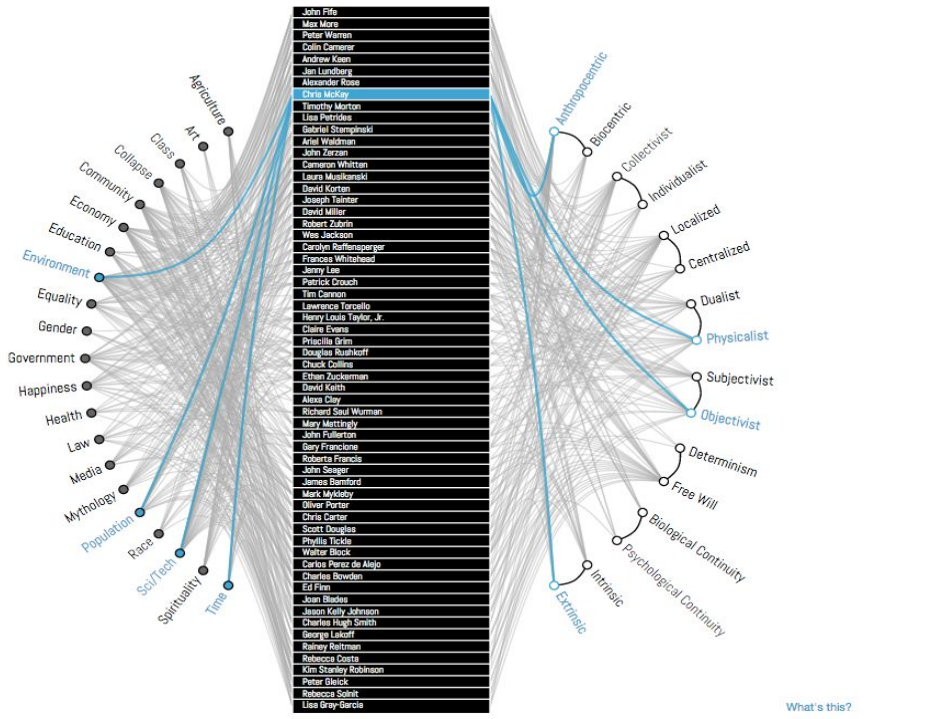
(http://www.findtheconversation.com/concept-map/)

After going through a number of D3 designs online with an intention to get some inspiration, this particular concept map really inspired us to build over its design as it brings out the one to many relationship between the fields in the tiles and the fields in the encompassing circle. This instantly struck a chord with us and gave us the idea that we could show the one to many relation between the genres and actors who have won awards for movies in those genres very well.

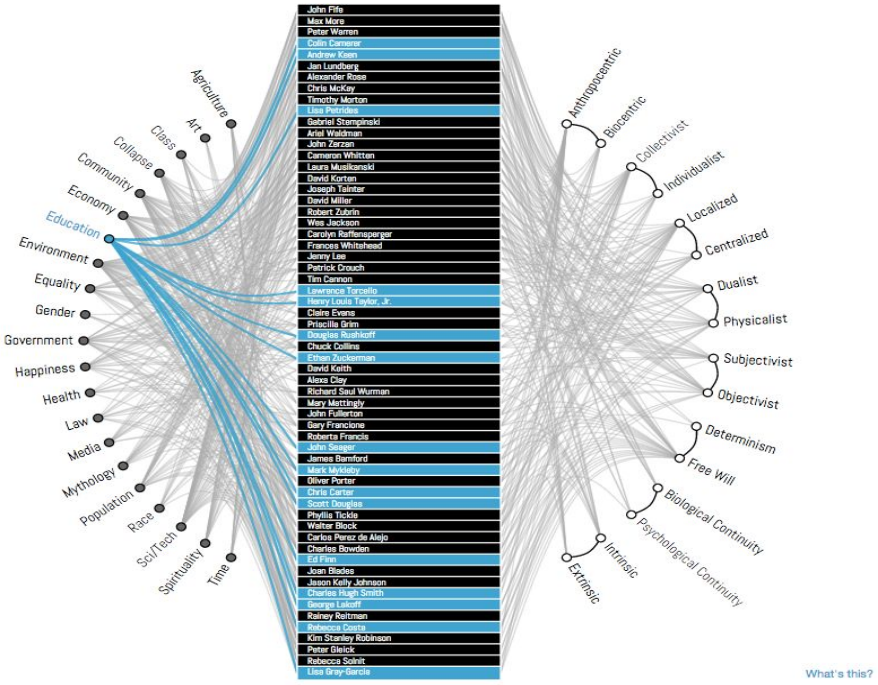
It took us a few iterations to decide which field between the actor name and the genre goes in the center versus in the surrounding circle.



Default view: The Conversation concept map with no highlighting



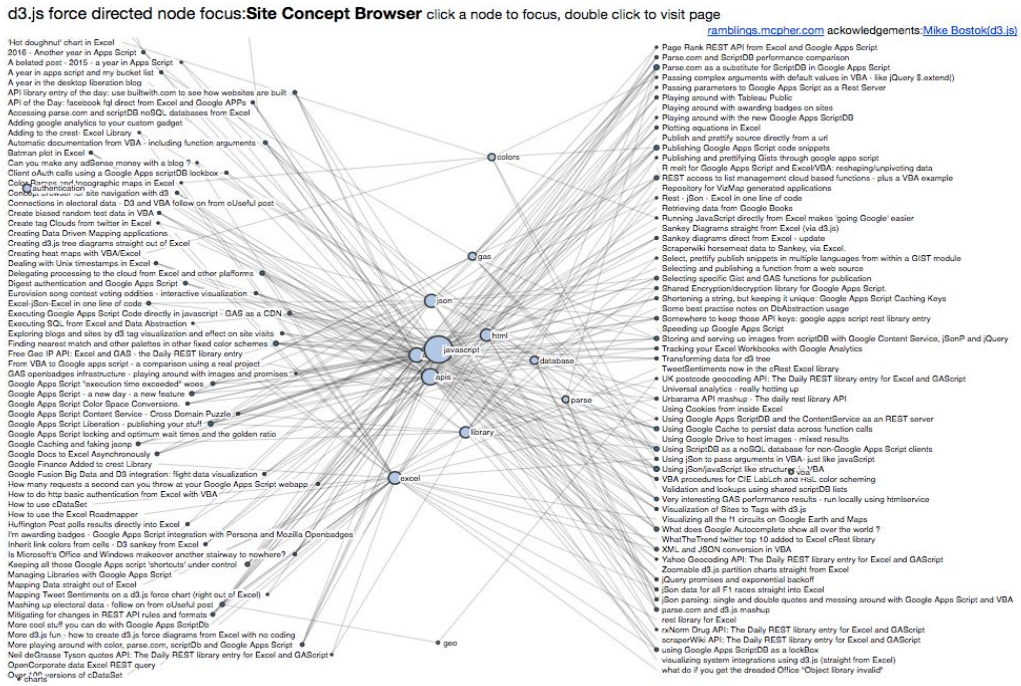
On mouse hover upon the fields in the center list: Relevant nodes are highlighted



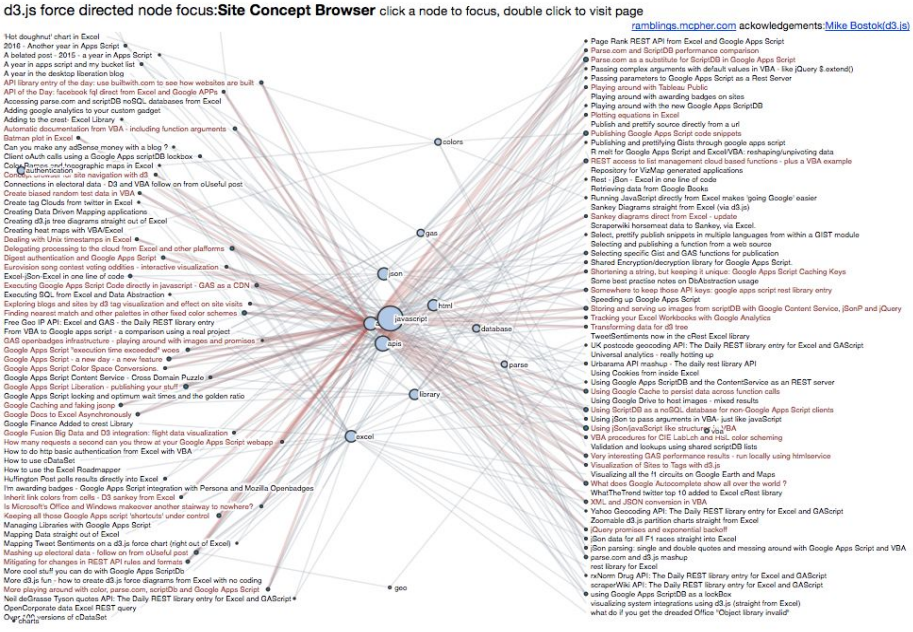
On mouse hover upon the nodes: Relevant fields in the center list are highlighted

ii. Related work 2: *Site Concept Browser* (http://xliberation.com/googlecharts/d3concept.html)

The particular design again reinforced our idea based off of the Conversation concept map and helped us confirm our idea inspired from the related work 1 above. We really liked the highlighting on mouse hover that happens and this helped us understand how this highlighting feature will help the users to better focus on the individual relationships. Although this particular design does more on the focussing part once the user clicks on a particular node but we eventually decided against it for our design as the requirements of our end goal didn't really match that. The main aim of this visualization was to just focus on the 2 parameters- genre and movie actor who won awards. Focusing on these 2 parameters, would help the user on getting the desired information in a crisper way.



Default view: Site concept browser with no highlighting



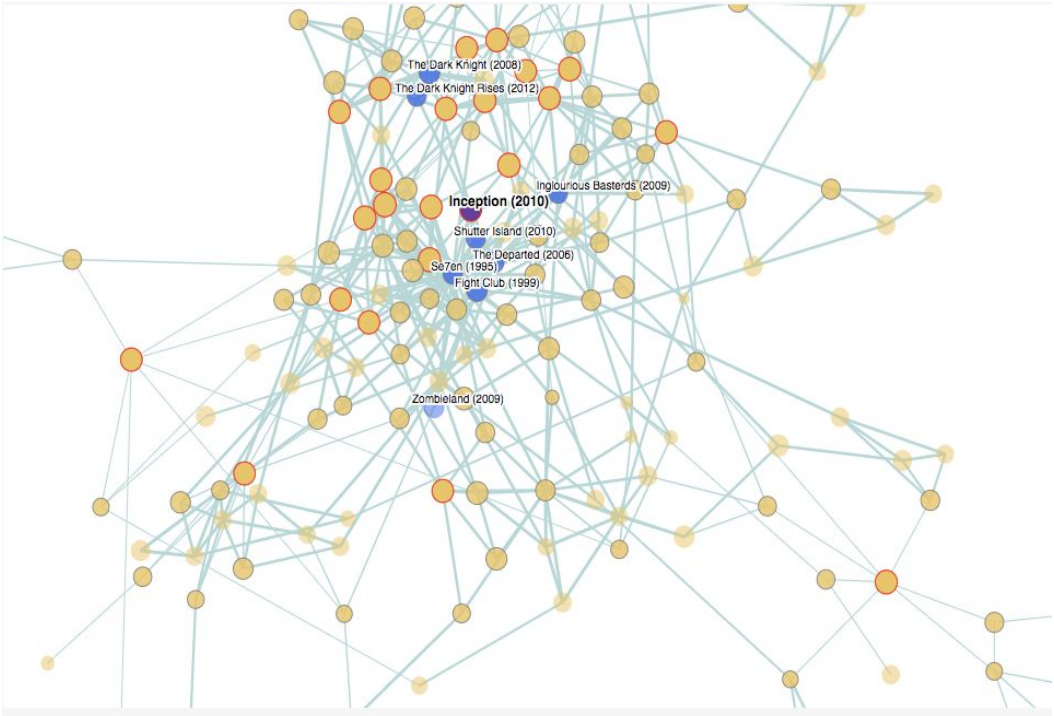
On mouse hover: Site concept browser with highlighting

iii. Related work 3: *The PopCha Movie Network*

(http://bl.ocks.org/paulovn/raw/9686202/)

This visualization is about movies itself and that's how it caught our attention. We could relate to this visualization as its topic was movies. We found the tooltip to be helpful that shows up on mouse hover on the nodes. Since this visualization is catering to a lot of data with no text/context upfront, upon hovering having the tooltip show up with the contextual information is helpful. This feature inspired us to include a tooltip for this visualization so that the user is presented with additional information like the names of the movies. This piece adds an important angle to the story telling for the user to relate to the visualization in a better way.

Default view: The PopCha Movie Network without any text/context



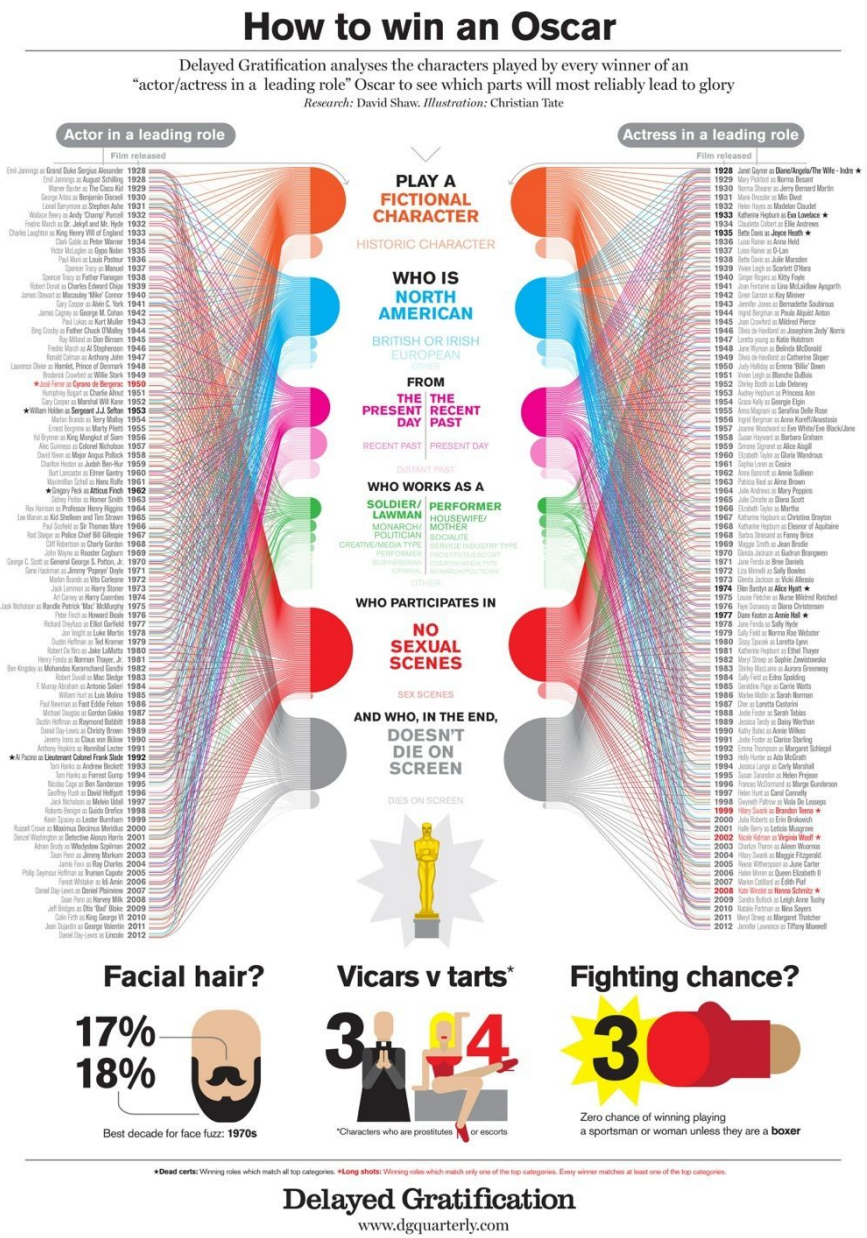
On mouse hover: The PopCha Movie Network with related information in the tooltip

Inspirations for Visualization 3

The visualizations below helped us in designing the various sketches that we used for the initial and final design. There were also takeaways from strength of certain visualization and also the weaknesses of the others.

i. Related work 1: *How to win an oscar*

http://www.rsvlts.com/2013/10/30/how-to-win-an-oscar-infographic/

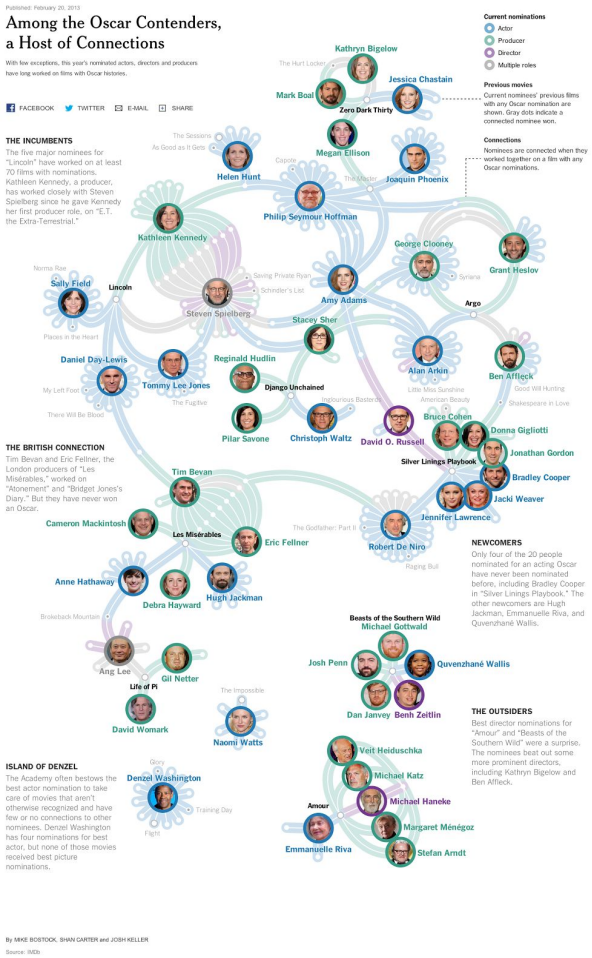


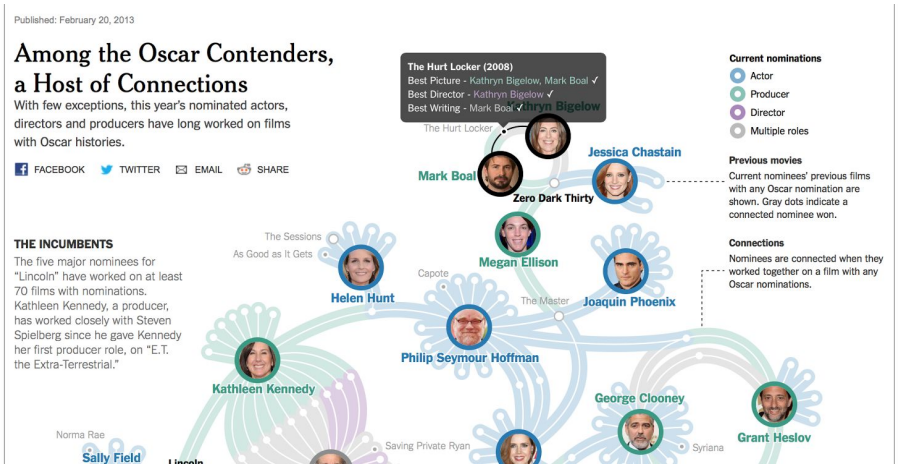
Another inspiration for our work include this infographic which shows how to win an oscar. This used a number of principles of Information Visualization like using colors to differentiate the facts and hues of colors within a fact to differentiate the sub-facts. It also makes good use of icons to show the facts which stand out or are interesting. They have also used some extra symbols near the name to show long shots and dead certs. It also shows lot of other information like the year and character name. However, the visualization is not interactive and might be difficult to search for an actor unless the user knows the year range in which the actor won an award. It also lacks few important information like director name and movie name. It might be difficult for the user to engage with this visualization unless they know more about the specific character in the

movie. Maybe the targeted users for this infographic are avid movie goers who associate themselves with the movie's characters. Some inspiration from this infographic are to make the visualization interactive by using a search or make it easier for users to find information by listing it alphabetically. Use of lines to show links or relationships and to tone down the data to a small subset. This visualization uses an extended dataset from what we have to explore patterns within the award winning actors and their character in the movie which contributed to the award.

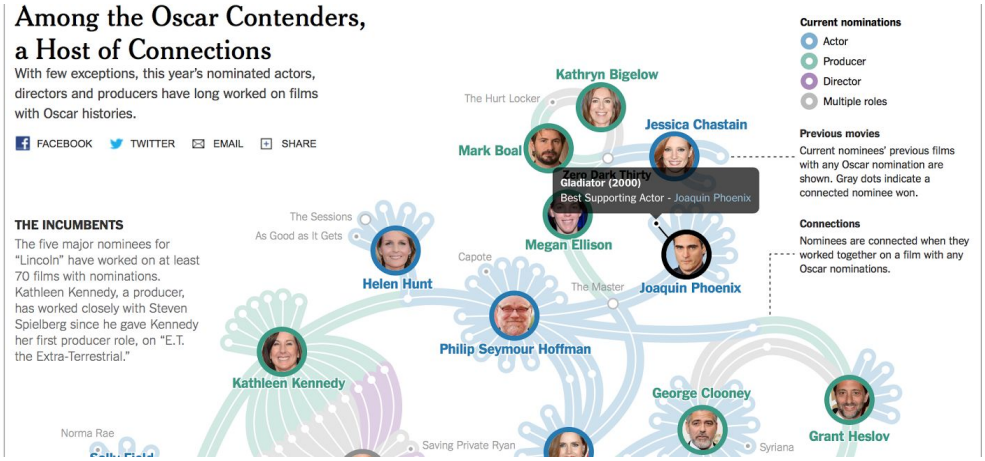
ii. Related work 2: *Connections between the oscar contenders*

http://www.nytimes.com/interactive/2013/02/20/movies/among-the-oscar-contenders-a-host-of-connections.html?_r=0

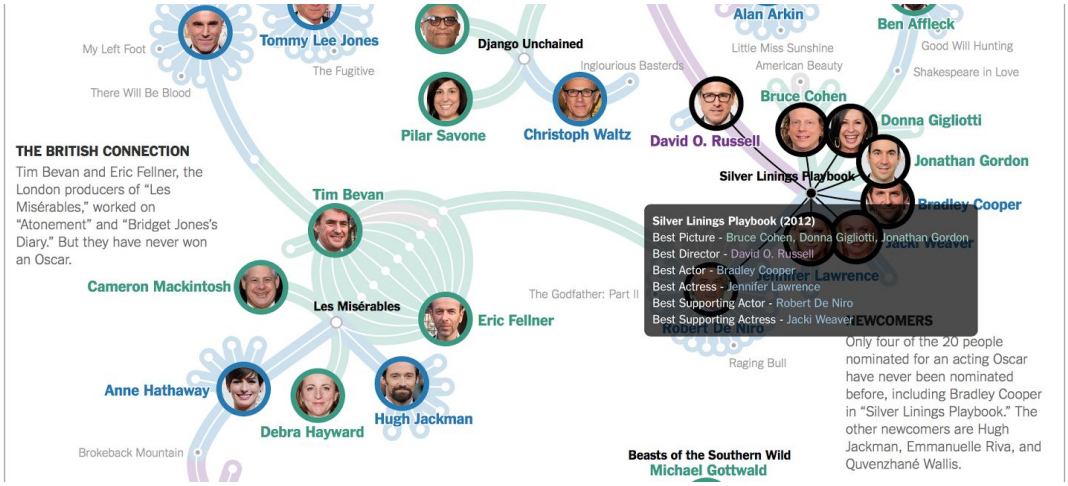




More information about the connecting movie



More information about the nominated actor

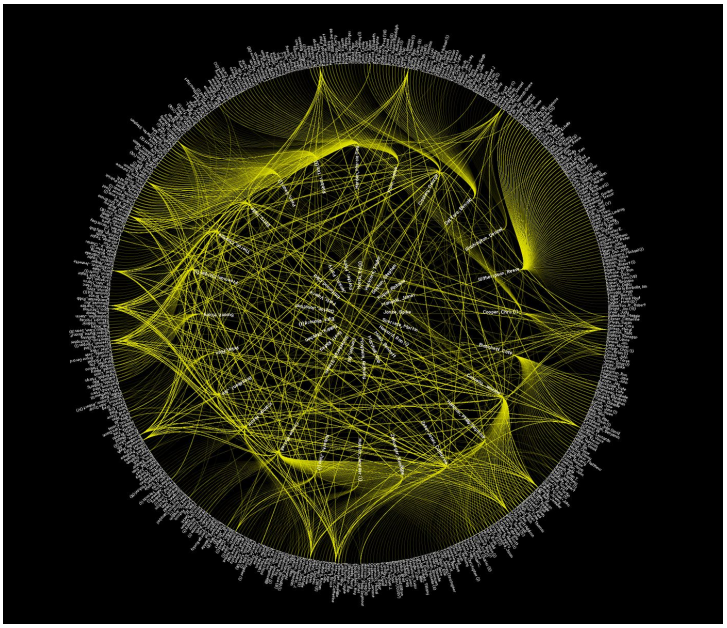


More information about the nominated movie

We obtained a lot of inspiration from this visualization which shows complex relationships between the actors, directors and producers including the movies acted, awards won etc., for the current nominees in the award ceremony. The visualization is also interactive making the user to focus on the information that the user is most interested in. It gives information about how the nominated best picture's producer connect to each other, how the nominated directors connect to other directors/actors/producers etc., It differentiates between nominees and oscars winners by a tick mark. One of the main takeaway from this visualization is that complex relationships can be portrayed well when the dataset is limited. And there are many past and current relationships between the actors and directors and actors and other actors which could be overwhelming if it involves a huge amount of data. This inspired us to narrow down our dataset to a few years and only to actors who have won awards and to focus on one relationship i.e. between actors and directors. The color coding, using tooltip to show movie, award information and links between the relationship were other takeaways. This visualization uses a subset of our data along with some external information like their origin to portray relationships between the actors, directors and producers nominated for the award in a particular year.

iii. Related work 3: *Directors and award winning actors*

http://www.pitchinteractive.com/infovis/abstract.html?utm_source=datafloq&utm_medium=ref&utm_campaign=datafloq



Another visualization which captured our attention is the above showing the relationship between the award winning actors and the directors and other actors that they have

worked with (the period is not apparent/not mentioned explicitly). A subset of this visualization is what we ended up doing but with focus on award winning actors over several years. This visualization is quite complicated and difficult to read. The center ring shows the award winning actors, the next ring shows the directors they worked with and the outer ring shows the other actors they worked with. It is a good visualization but the lack of interactions and additional information about the movie is a disadvantage. This inspired us to settle in with the final design for our third visualization with the relationship between award winning actors and directors they worked with and add the vice versa relationship of award winning actors who worked with that director. We also made sure that our visualization is interactive and can handle searches. We also reduced the dataset to show a less complex view.

3. Description of the visualizations

Visualization 1*:*

Initial design:

Our main idea was to create a design through charts that allows the user to explore our the huge amount of data which is made by web scraping IMDB and by combining two data sets – IMDB movies and awards to look for any patterns/trends or any particular insights that throw some light at some of the research questions that we were trying to explore like :

1) There is a popular notion that although academy award winning movies tend to have great story, technical and creative expertise and most of them may even be critically acclaimed, but many people perceive these movies as movies that are based on complicated and dense topics and ideas that sometimes use very poetic or complex way of sending a message across to their audience (i.e through metaphors or by asking the audience to read between the lines) often causing a lot of people to stay away from such movies –questioning the fact whether academy award winning movies are in fact popular with general audience.

2) This further led to a question that, in the recent years, do critical score and popularity of a movie go hand in hand.

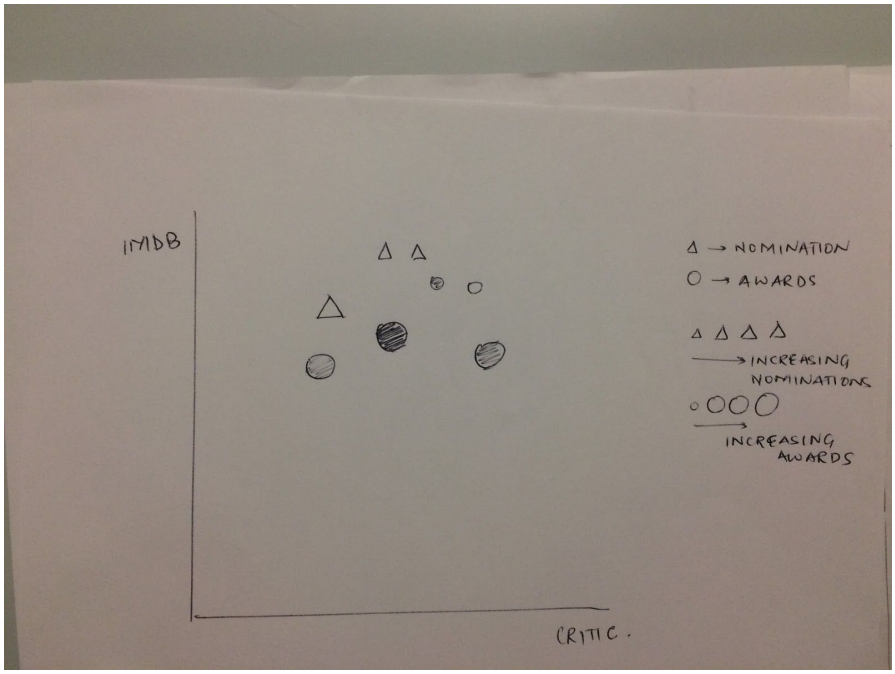
3) We have always seen big movies like Titanic, Lord of the rings which have big budgets win record breaking awards in a single night, it would be interesting to find out whether big budgets have consistently resulted in awards or whether having budget does not necessarily guarantee that you will win an award.

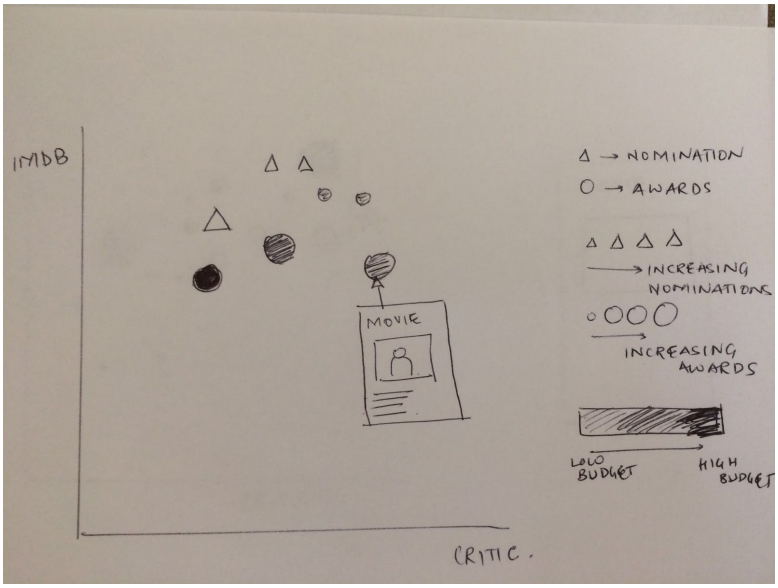
4) Do award winning movies often belong to a particular genre.

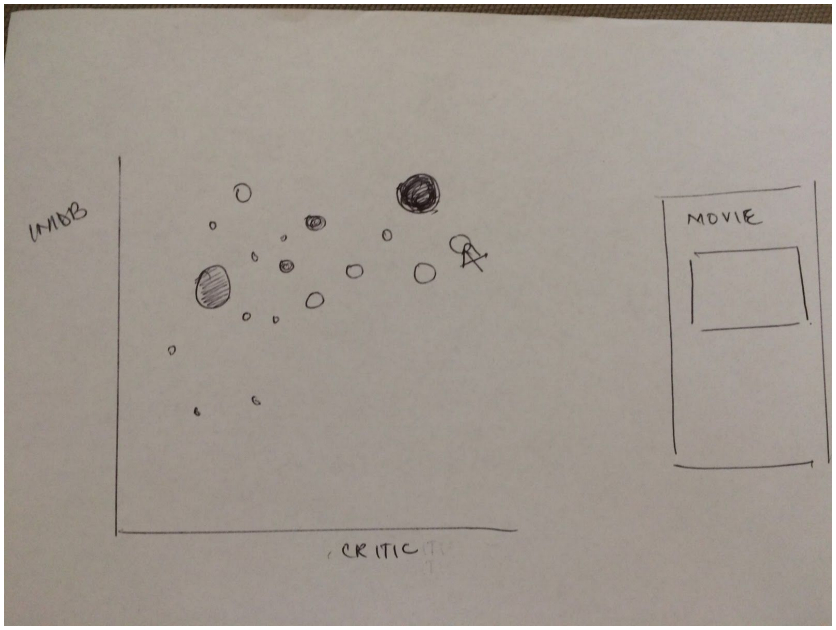
In order to come up with a way of showing this data, we came up with a lot of initial designs and tableau explorations. The data that we collected allowed us to plot movie data based on many factors like - year released, the type of awards won, nominations it received, budget, IMDB score, critic score, facebook like for various actors, directors and movies and more.

We initially considered making a dashboard through tableau as we were thinking about incorporating charts. It sounded like a nice idea to allow the user to do exploratory analysis by combining charts on various parameters. Having different charts in a dashboard would allow us to apply the heuristic principle of not overloading memory and inundate the user with a lot of information in a single chart.

However, having information in different charts would take away the opportunity to see how trends between other bits of information represented in neighboring charts interacted with each other. Additionally, having separate charts, would keep some bits of information separate, which would make it hard to discern any trends that would otherwise be apparent had the information be displayed in a single chart.







We then toyed with idea of plotting data for each of these factors based on yearly data that we had to present yearly trends. However, plotting a lot of information on yearly scale meant a lot of information that needed to be grouped or changed to make sure it is visible in one view. However, since we wanted to keep the data used across all the visualizations to be consistent, we decided not consider yearly data and decided to restrict data to only last 20 years, so that we have fewer movies, actors and directors that we could easily plot on the other two visualizations without them looking chaotic. Moreover, in order to answer research questions it was not essential to have yearly data in our charts. It was more important to have data like IMDB score, Critic score, budget, genres and academy awards on the chart.

Thus, we scoped down to these parameters and decided to incorporate a feature that allowed the user to search based on actors, movies, directors etc just like we saw earlier in another inspiring info visualization (shown in related works #1) . Applying Few's guidelines and inspired by the way information was represented through color shape, size and position in the chart – " The Best in Show: Ultimate data –dog" , we decided to visualize movies based on the four main parameters (critic score, imdb score, awards/nominations and budget and genre) that would help the viewer to answer the research questions, on the same chart and have few features like actors, directors, movie names to be used to filter data. We then tested few sample designs to experiment with idea of incorporating may be shapes, sizes and colors to represent information for nominations, awards or genres.

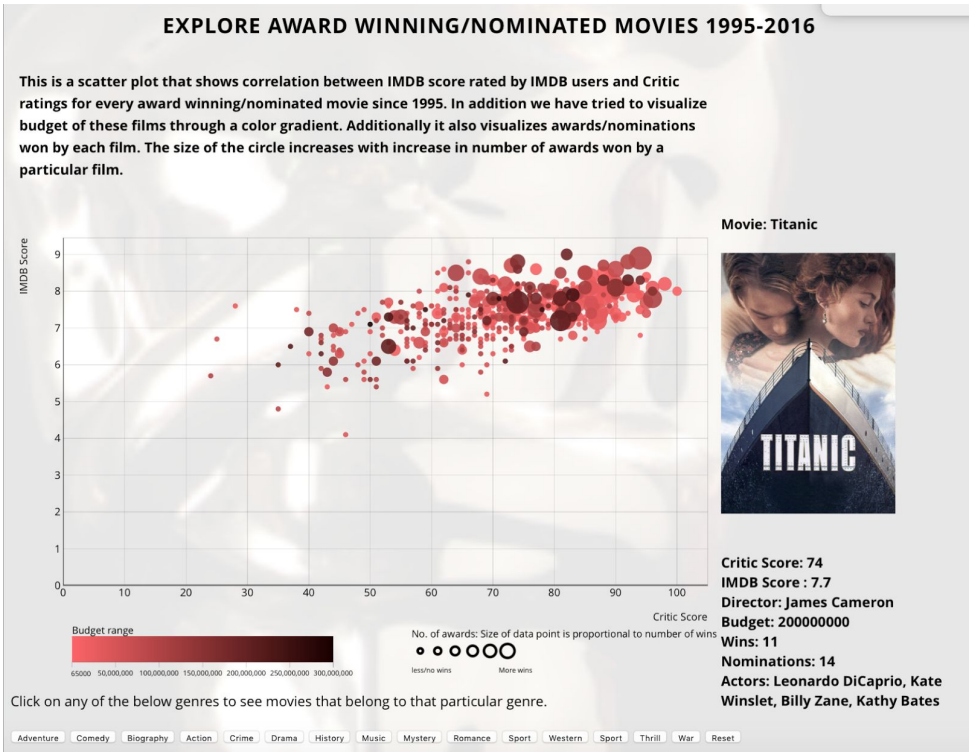
Final Design:

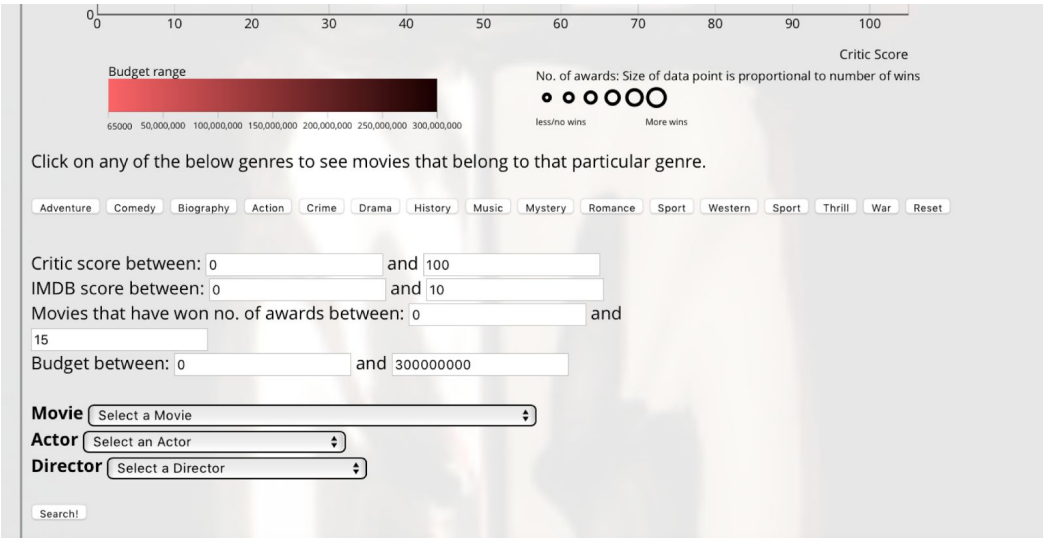
Code references:

http://bl.ocks.org/peterssonjonas/4a0e7cb8d23231243e0e

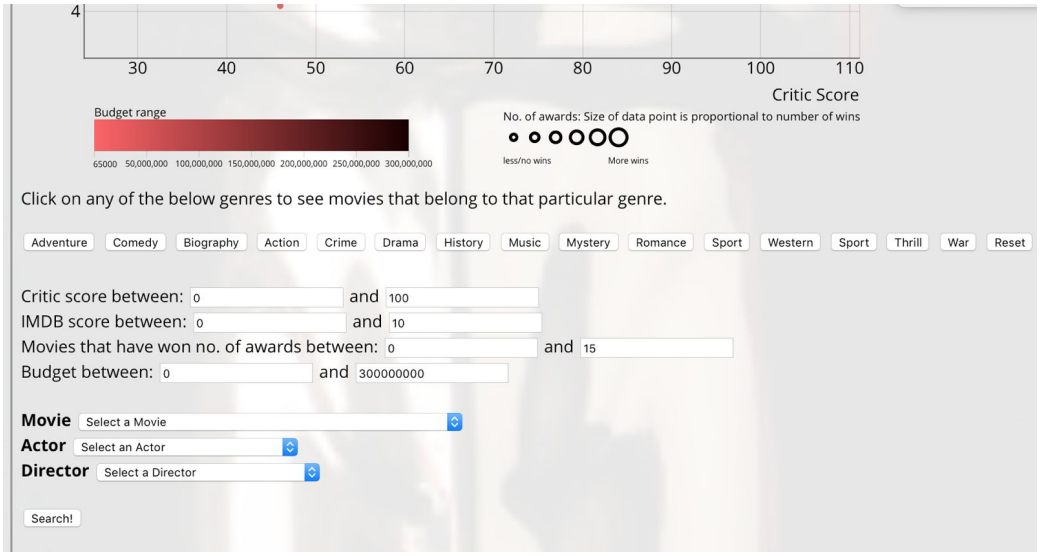
http://stackoverflow.com/questions/28922246/dynamic-filtering-in-d3-with-html-input

After receiving some feedback from user testing, we then decided to visualize the critic score and IMDB score on the axis, as the correlation between these two parameters would be most clearly conveyed only if they are directly plotted on a graph and hence we decided that the awards should be represented as size (the size of the circle is directly proportional to the number of awards won) as there were fewer and more discrete values for awards and the awards information being more salient than budget, since the idea is to look for a trend that signals to a common theme that exists between award winning movies and also the number of awards won with other parameters. Similarly, we then decided to represent budget by using a gradient in colors.

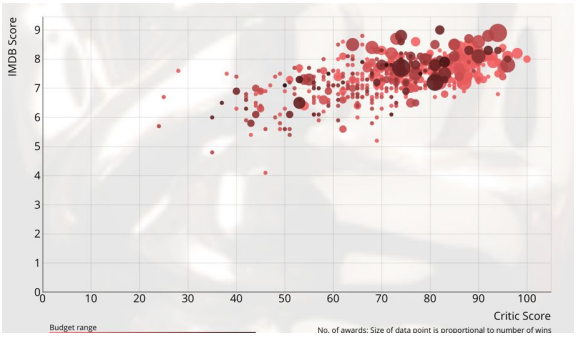


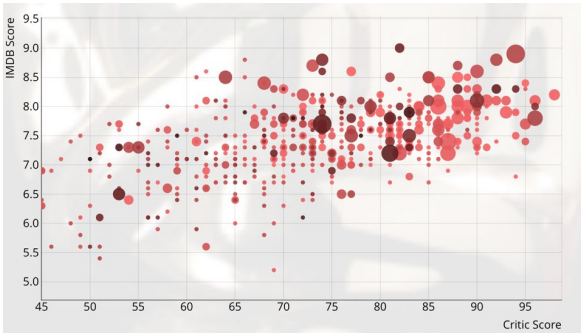


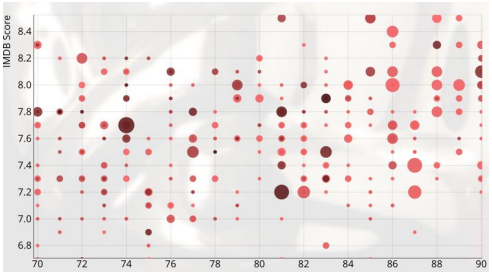
We initially had features to do a search only based on movie titles, actors and directors and to see how the chart changed for each genre. However, after receiving feedback on our design, we decided to incorporate more features like filtering data based on critic score, IMDB score, number of awards and Budget. This way now users can dissect and restrict data to see any way they like.



We also decided to use D3 here instead of Tableau, because we wanted to give user the ability to zoom on any portion of the chart and we wanted the axes to scale accordingly. This was a an essential feature because it let the user navigate and only concentrate on the data they were interested in looking. This becomes an important feature, especially given the fact that many movies have IMDB score around 7 and more and thus tend to crowd and overlap over each other in a specific region. Having the ability to zoom in would allow the user to clearly see what's happening in a specific region in the chart and thus capture any trends that might have otherwise been lost in a sea of overlapping data points.

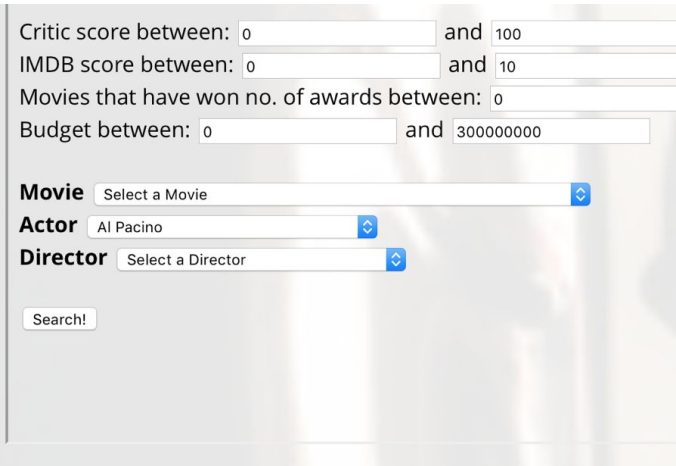




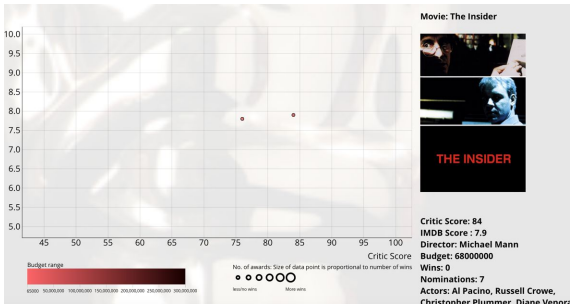


The users can filter data as they like based on movies, actors or directors – all of which are drop down menus which are alphabetically arranged. Since we have limited the movies that we are showing in plot only to the ones that have won awards or were nominated, the result of selecting a particular actor /director will only give us the movies for which that particular artist has won awards or has been nominated between 1995 to 2016.

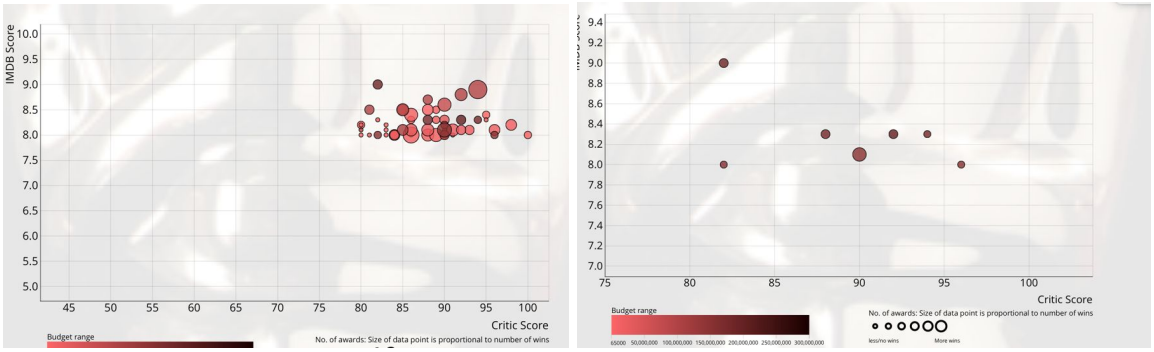
Eg: Search for movies actor “Al Pacino” has won awards for or been nominated between 1995-2016.



one movie selected – “THe Insider”

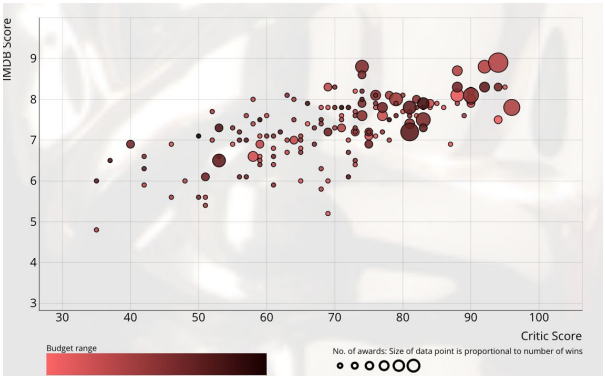


Similarly, data can also be searched based on entering a range of awards, critic score , iMDB score etc. The below picture shows the data for movies that have IMDB score more than 8 and critic score beyond 80. The picture beside it lets us see data further restricted to a budget of a film that lies between $150 Million dollars to $ 300 Million Dollars and zooming in this data allows us to see this information clearly.

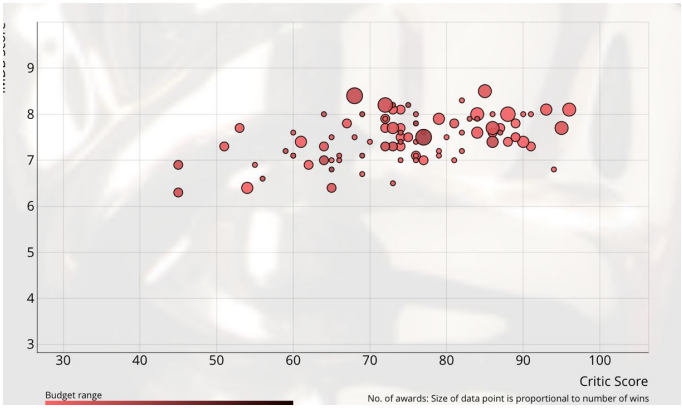


We can also see how all the data changes for each genre.

Adventure: (Interesting insight: most movies in the adventure genre have high budget).



Biography: This genre has mostly lower/average budget movies



We can also use the rest button to reset the graph and get all the data points back on the graph.

We also used D3 over Tableau to allow the information about the movies to displayed beside the chart whenever mouse hovers over data points.

Some interesting insights and answers to the questions we raised initially:

1) From our graph, we see that there are more circles with bigger in size that are more accumulated towards the far end and towards the top of the graph, signifying that in fact contrary to these popular notion that we discussed before, academy award winning movies are in fact not only popular with the critics but also popular with the general audience.

2) Similarly answering the second question, given how there is more crowding of data points towards top right of the chart, one can say that more critic score and imdb score(taken as a measure of popularity) do go hand in hand.

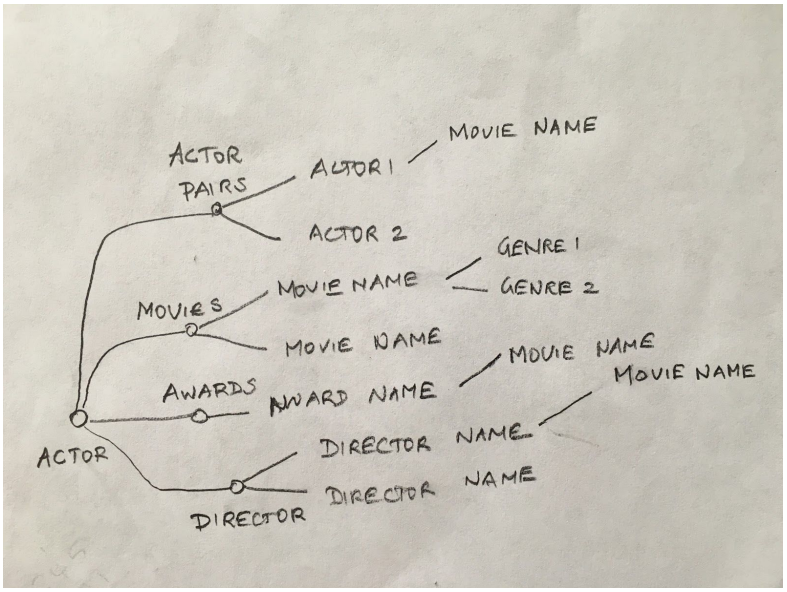
3) Given how there are few big dark points that are found in the center of the plot and more dark points that are smaller in size , signifies that movies with higher budget have not always won more academy awards. In fact the movies that lie far right of the chart and that are bigger in size tend to be lower in budget or have average budget.

4) Most movies as that have won awards belong to the genre of drama.

After some usability we got a good response and people loved to explore different aspects of the chart. However, some improvements that we found that need to be made are like adding more features to make certain movies invisible features invisible. Add more brushing and linking features that could apply to all designs. For example, being able to see how an actor's movies have performed in the first visualization and get more details about him through the second and third visualization. As of now this can be done but the filtering still needs to be performed manually on all visualizations.

Initial idea (visualization 2 and 3 as one)*:* We actually started off with the visualization above opening up a network to show more information like awards received, movies acted, directors acted with, actors paired with. But our survey showed that people did not care much about actor pairing but cared about director pairing and movies. Also, the prototype below was little confusing to the users as they could not associate the director, award and movie name together. We also thought about introducing another level to the director name with movies and awards and users were confused about if those awards belonged to the director or the actor himself. Also, if an actor was also a director would the award still show up. Who were the other actors with whom this director worked with? - Incorporating this into the original design will change the focus from actor to the director and his pairings which could be confusing. This made us to realize that giving option to drill down deeper into aspects within a single visualization could be misleading given the depth of the dataset. So, we split the design into 2 parts with the first part concentrating on award winning actors and genres associated with them, answering questions like "Which genre should an actor work in to get an award?"

and “Does a particular award winning actor choose a movie of certain genre?” .The second part focussing on director actor pairs of award winning actors to answer the question, “Do some directors have the history of producing movies where the actors win an award?”



First idea

Visualization 2*: A concept map linking movie genres and award winning actors*

This concept map focuses on 2 parameters- the movie genre and the award winning actor.

There are in order 15 movie genres for which we have the available data in the IMDB dataset. A particular movie can have multiple genres usually ranging between 3 to 8.

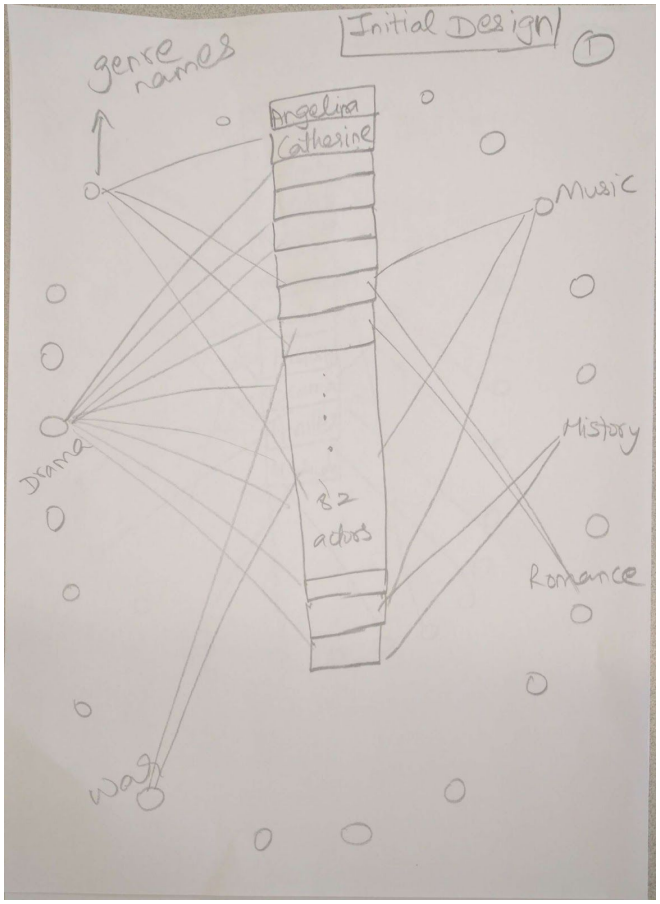
There are about 82 actors who have won an academy award (1 or more) in the awards dataset. We have included all these genres and all the award winning actors in this visualization. Since the number of actors is much more than the number of genres, the design looked more appealing when we placed the actors in a circular fashion surrounding the list of genres in the center. We had first tried the design in the opposite fashion with the 82 actors placed in a list in the center and the 15 genres surrounding it.

When we conducted usability tests on this design, we found that users were struggling to go through the entire list of 82 actors and were not able to find their favorite actor easily and quickly- some users would just give up halfway through and just hover randomly on the visualization and not get the experience that we had envisioned.

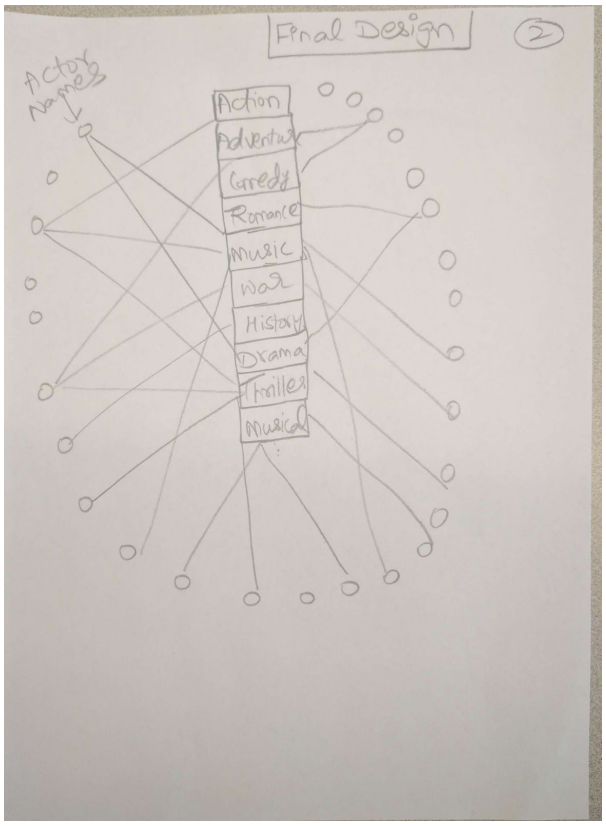
This compelled us to try out the alternate placement of the 2 parameters and usability tests showed that it worked wonderfully.

Also, since we want to narrow down the data, we decided to concentrate on just 4 categories of awards- Best actor in a leading role, Best actress in a leading role, Best actor in a supporting role and Best actress in a supporting role.

Paper prototype for initial design-



Paper prototype for final design-

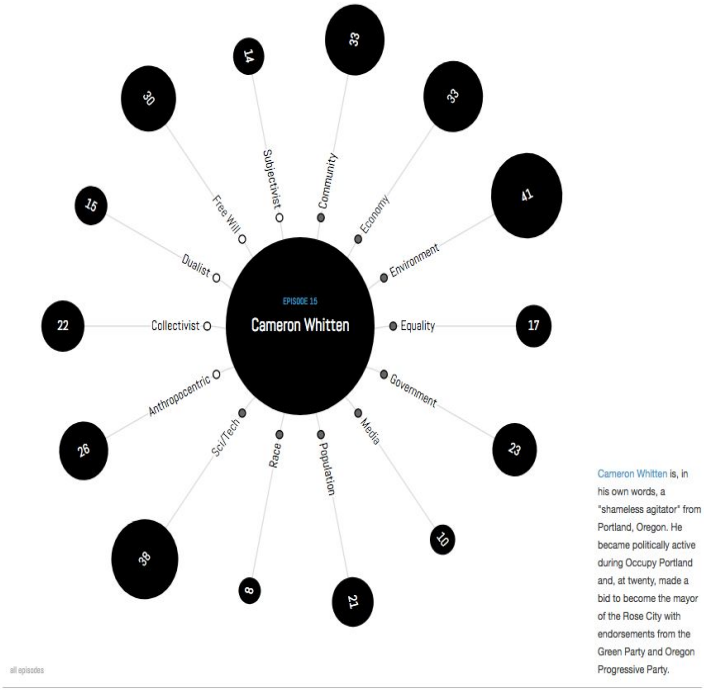


Some additional design considerations-

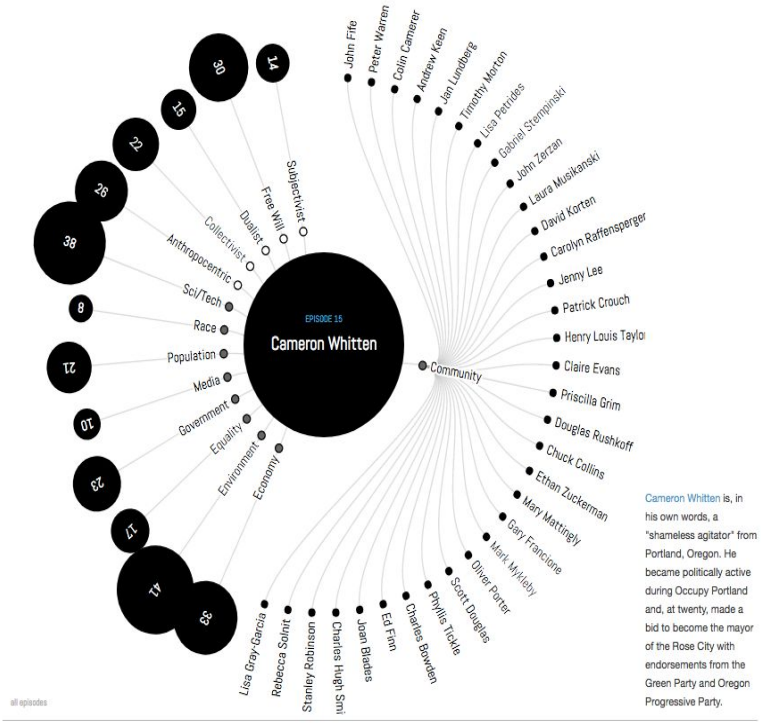
- *Tooltip*- when you hover on the name of an actor, you can see the name of the actor, the name of the movie, the award that the actor has won and the genre of the movie. This is basically included so that the user can clearly know all the related information at once and if an actor has won awards for multiple movies, the tooltip shows all the above information for all of the movies, thus giving the user the complete information.
- *Alphabetical order of the actor names*- In the first iteration, we had placed the names of the actors in outer circle in the order the names were available in the dataset but upon conducting usability testing, we observed that the users were finding it difficult to spot the name of the actor they had in mind and were taking a lot of time going through all the 82 names and sometimes would just skip the name and go through the 82 names all over again. We then figured out that this

problem could be overcome by placing the actor names in an alphabetical order of first names so that finding an actor name becomes easier and more intuitive.

- *Color scheme for the highlights*- To match the overall color palette, we kept the color of the highlights that happen upon mouse hover in sync to give a more consistent look and feel to it.
- *Not going ahead with network graph*- The concept map (related work 1) online had additional features of clicking upon a node that further drills down the data by one level as shown below-



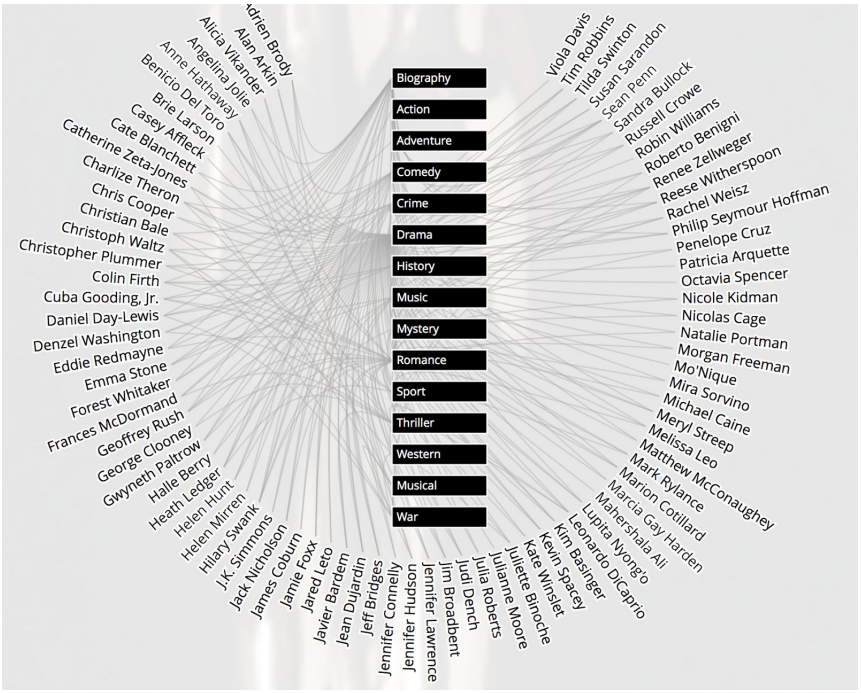
Upon clicking a node on the default view



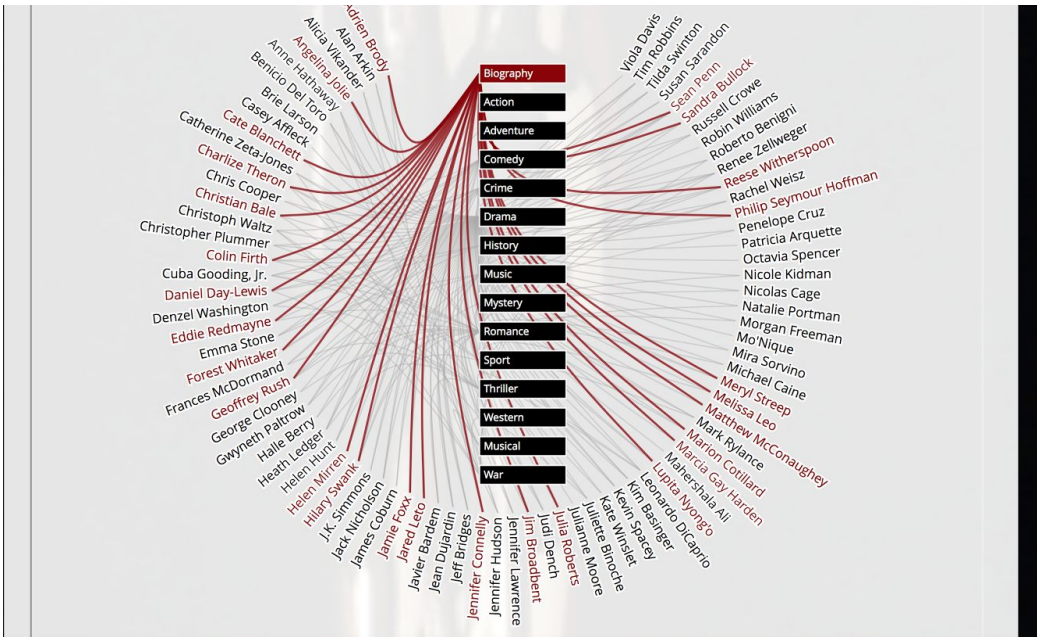
Upon further clicking a node in the view above

The reason we did not include this drilling down feature is because we wanted to focus on the particular task of giving the user the ability to find out the movie genre and award winning actor relationships. We had earlier included a level of drilling down in our first iteration but somehow we felt we are not conveying any additional information that way and drilling down reiterates the same information that we were displaying in visualization 1 only displayed in a different way. We found this to be redundant and a digression from the main point and focus that were trying to make in this visualization.

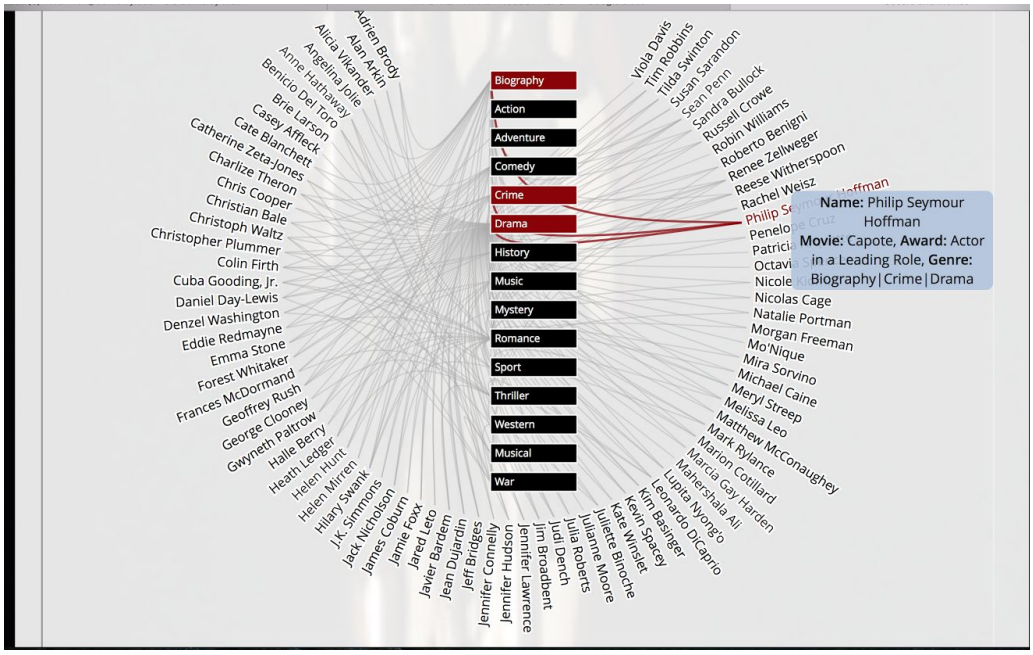
Although drilling down increased the interactivity with the visualization, the end goal that we wanted to achieve in this visualization was being accomplished without drilling down. So this was a tradeoff that we decided to make. The final design is as follows:



Default view



Hover over genre

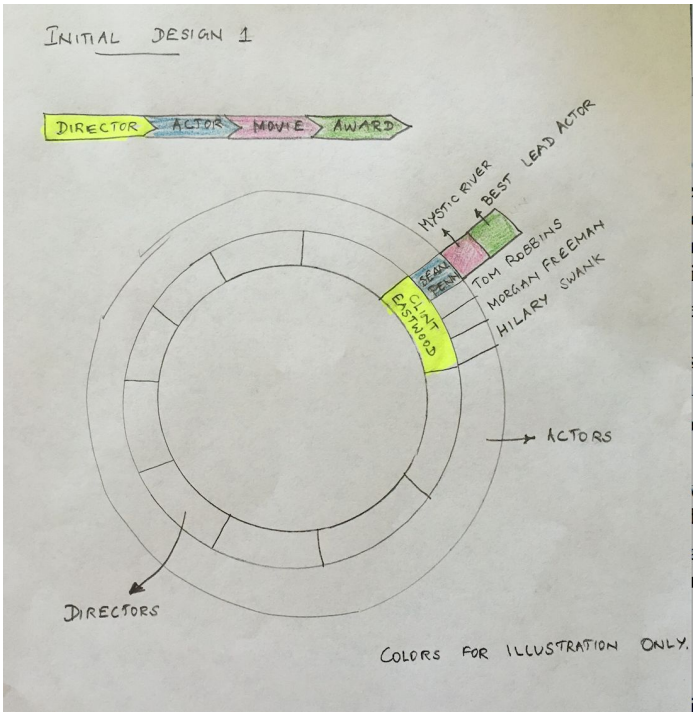


Hover over actor

Visualization 3*: A chord diagram linking the director to the award winning actors and vice versa.*

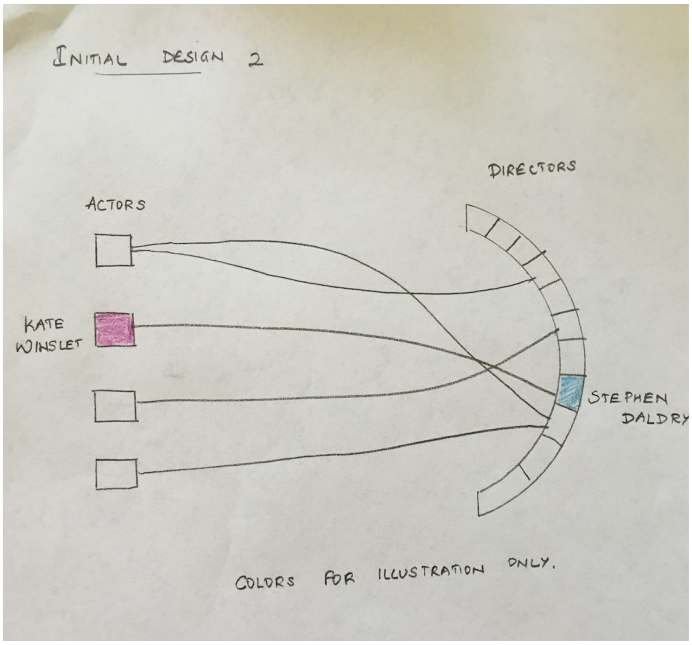
Prototypes:

We started with prototypes which we felt could work well for the director-award winning actor pairing that we wanted to focus on. We started with a hierarchical design in a chord diagram -Initial design 1. On choosing a director, the award winning actors who worked with the director are shown from which the movie and award can be determined. The drawback of this design was that if users were interested in knowing about all the directors that an actor has worked with, this design did not provide a way to get that information. Also, given the huge number of directors and actors showing arc segments would not be as pleasing and actors might be repeated if they had worked with more than one director.



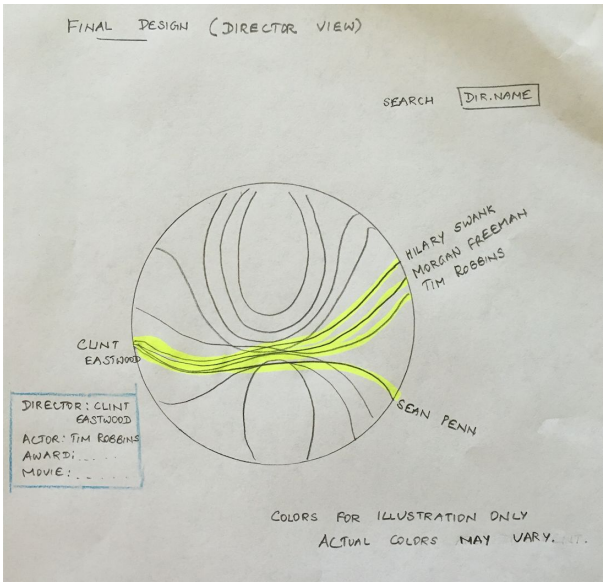
Initial design 1

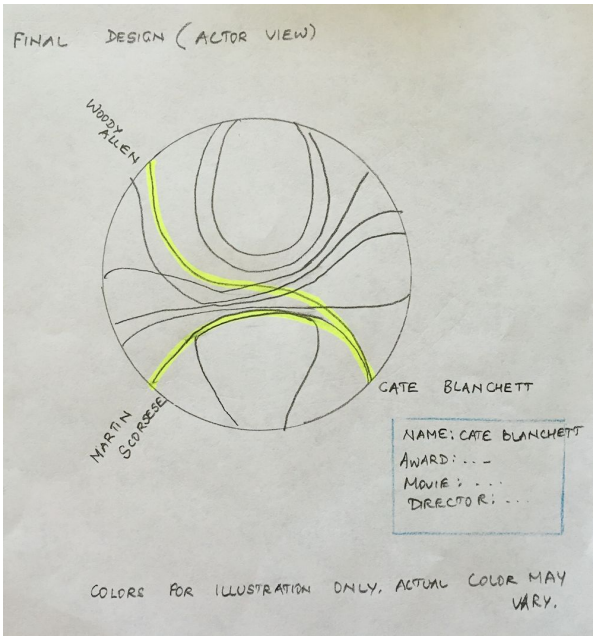
For the second design (Initial design 2), we considered a plot line visualization where we intended to depict the actor by a picture and associating them with the director they had worked with. But displaying pictures for more than 40 actors and directors would disorient the users and the size and quality of the image would be lost in the process. So we decided to skip the pictures and just go with names of the actors and directors. This design allowed interactions in both the ways, where the award winning actors associated with a director could be traced as well as vice versa. We tweaked the design a bit and ended up with a final design which was a mixture of both the initial designs.



Initial design 2

Final design: The final design used a chord diagram but without arc segments. The actors and directors could be separated on either side of the arc and links between the directors and actors can be traced to understand the relationships between them. We chose to display the award winning actors along with directors whom they worked with in that movie. This answers 2 critical questions:”Which directors tend to produce movies where the actors win an award and who are the actors?” and “Who are the directors that an award winning actor has worked with?”

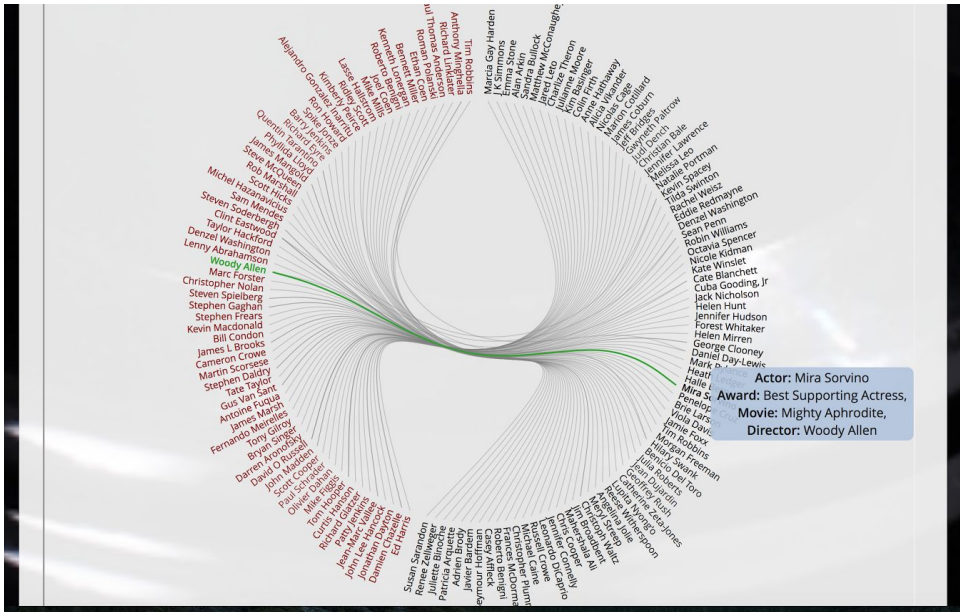




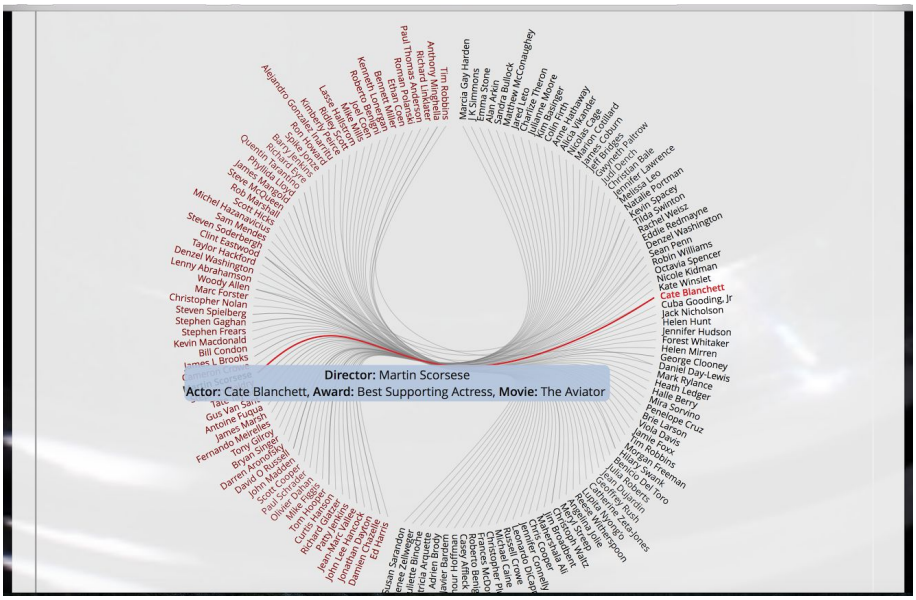
Final design

Code and add-ons: We found the code from https://bl.ocks.org/mbostock/7607999 and used it for our data with changes to the design. We liked how 2 relationships could be portrayed by the design but did not like that there was no distinction between the parent and child by color. So, we incorporated colors to differentiate between the director and actor. We chose the colors to be black and red to blend in with the overall theme of oscars and chose to highlight the links using red and green as in the original design as it worked well with the color scheme.

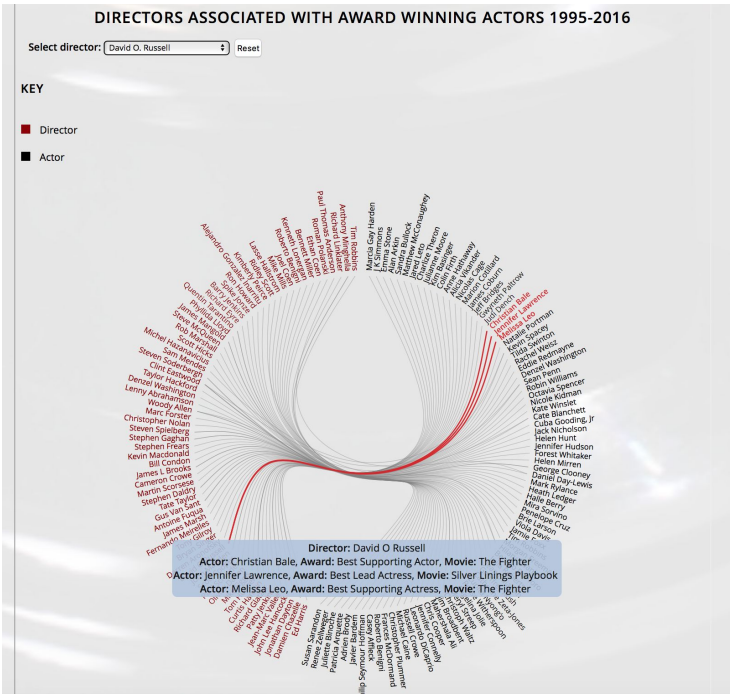
Also, looking for a particular actor/director was difficult with the given design. So, we added a drop down select to the design where users can choose a particular director and since we were more interested in the director we chose not to include the drop down select for the actor. We also included a tooltip to the design to give more information to the users about a particular director or actor relationship like the award name, movie and actor/director. The final visualization with the colors, drop-down search and tooltip are shown below.



Hover over actor



Hover over director



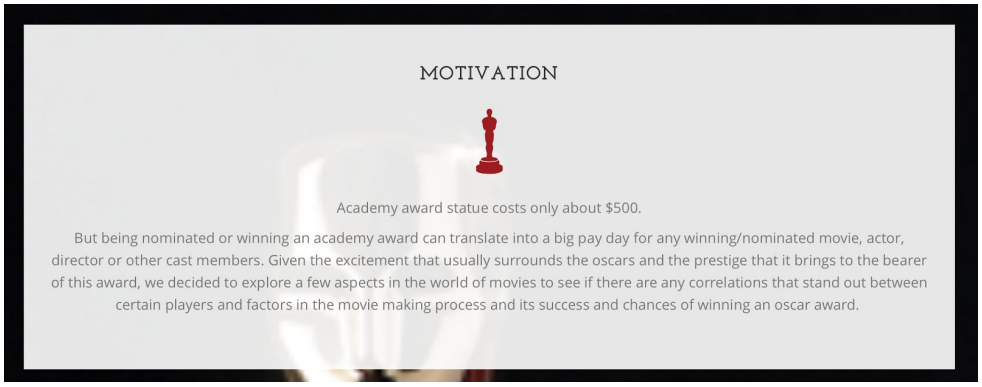
Select a director

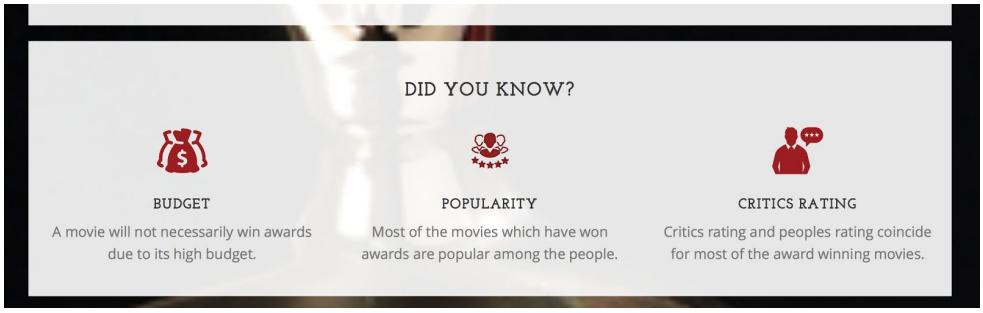
Finally, we conducted usability studies for this design and we received positive response for the design. The users liked how there was a 2-way relationship between the actors and directors. Some did express the inability to show award winning directors. We felt it could be incorporated using a symbol to show if the director himself

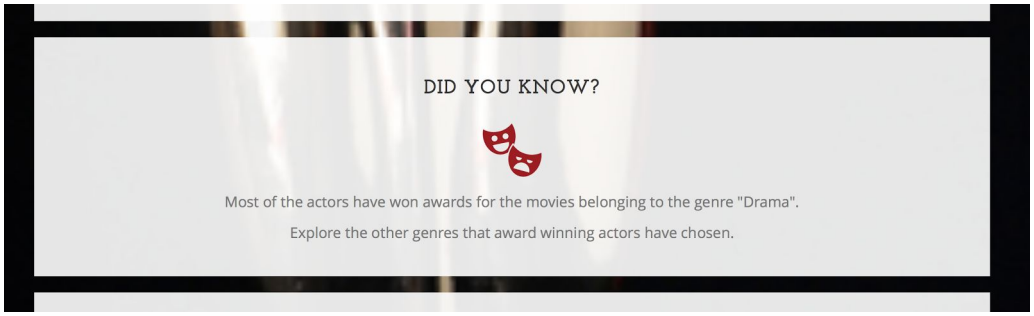
won an award and displaying more information about the award. This could be an improvement or additional information which could be added to the design.

Overall flow of design:

In order to tie the 3 designs together, we decided to write some facts and motivation for our visualization. We included less known facts along with infographics to kindle the interest of the user and to make them look for more facts that could be discovered through the use of our visualization. For example, we put up facts about the critic score and IMDB score as a fact using infographic which would make the users to want to know more about the other interesting facts. We first decided to use external facts and numbers like how much a movie earns before and after nominations but after usability studies found the numbers to be distracting to the users and out of context of the visualization that we were trying to achieve. We chose to stick to the black and red theme for the entirety of the webpage to tie with the red and black theme of the oscars. We also tried incorporating gold color in the visualization to capture the oscar statue itself but the color was too light to capture the attention of the users given the bright nature of red and black in comparison. We used the template from https://blackrockdigital.github.io/startbootstrap-agency/, https://startbootstrap.com/template-overviews/clean-blog/ for our webpage and added icons from Gregor Cresnar, Thomas Helbig, Kokota, Iconoci, Tuk Tuk Design and John Melvin. Also a background of the oscar statue added more context to the webpage itself without overwhelming the overall design.







4. Data that were used to accomplish the goals

i. https://www.kaggle.com/theacademy/academy-awards has details like Year, Ceremony, Award, Winner and Film from 1927-2015.

ii. https://www.kaggle.com/deepmatrix/imdb-5000-movie-dataset - It contains 28 variables for 5043 movies and 4906 posters (998MB), spanning across 100 years in 66 countries. There are 2399 unique director names, and thousands of actors/actresses. Below are the 28 variables:

"movie_title" "color" "num_critic_for_reviews" "movie_facebook_likes" "duration" "director_name" "director_facebook_likes" "actor_3_name" "actor_3_facebook_likes" "actor_2_name" "actor_2_facebook_likes" "actor_1_name" "actor_1_facebook_likes" "gross" "genres" "num_voted_users" "cast_total_facebook_likes"

"facenumber_in_poster" "plot_keywords" "movie_imdb_link" "num_user_for_reviews" "language" "country" "content_rating" "budget" "title_year" "imdb_score" "aspect_ratio"

We also scrapped the movie posters and critics score from IMDB website for the movies that won an oscar or were nominated. We also collected the top actors and directors from the site as the data from Kaggle had many errors. We scoped the data down to years between 1995-2016 and focussed only on award winning and nominated movies, actors and directors.

5. Which tools were used to accomplish the goals

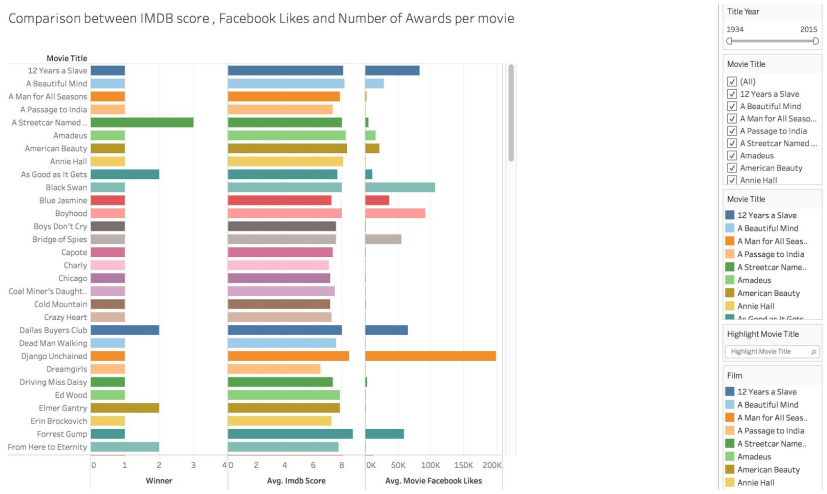
Tableau - We did some preliminary EDA using Tableau with the full dataset which showed us the complex relationships and the enormity of the data. This helped in narrowing down the dataset.

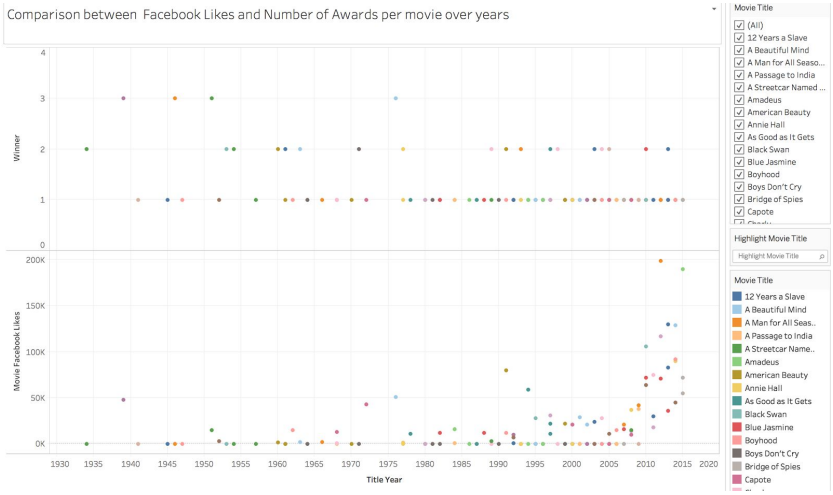
D3 - We used D3 to make all of our visualizations as it enabled us to customize the visualizations like highlighting certain values on selection, filtering of data and enabled adding posters to the movie information container. A few of these things are available in Tableau though d3 gave us enhanced visualization.

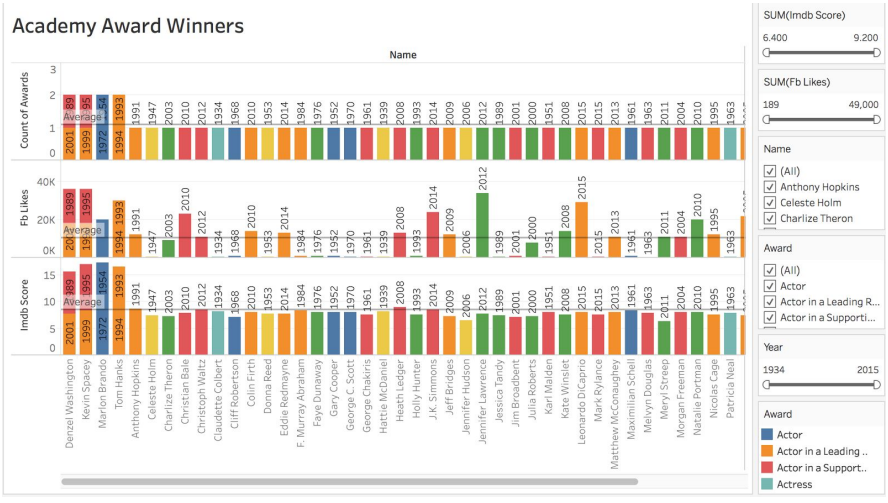
Illustrator - We used illustrator to change the icons to suit the color palette we had chosen to create a flow in the webpage and kindle interest in the users by giving some interesting facts. Icons were borrowed from Gregor Cresnar, Thomas Helbig, Kokota, Iconoci, Tuk Tuk Design and John Melvin.

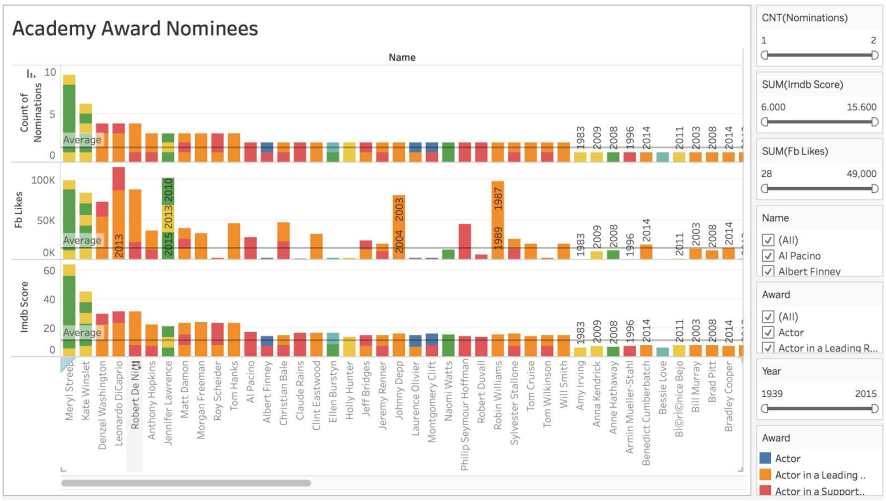
6. Steps that were required to accomplish goals

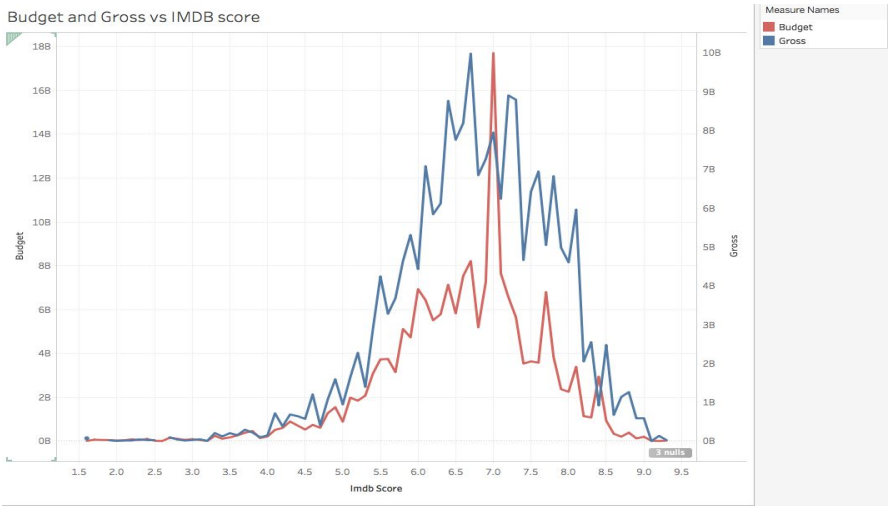
We combined two datasets on movies and looked at the data based on time, movie, actor, actress, director, awards etc., and plotted the trends in tableau initially to understand the data better. We then looked for anomalies and tried to explore the abnormalities in detail. Also, we had enormous amounts of data which could actually clutter the visualization if not chosen with caution. Given the number of movies, actors and directors, we felt the need to narrow it down to a few years of data and concentrate on movies that won an Oscar or were nominated. Incase of actors and directors, the list was still exhaustive and we decided to scope it down to just actors that won awards. This gave us a better model that could be visualized in interesting ways. Scoping down the data to the important facts required lot of EDA as shown below.

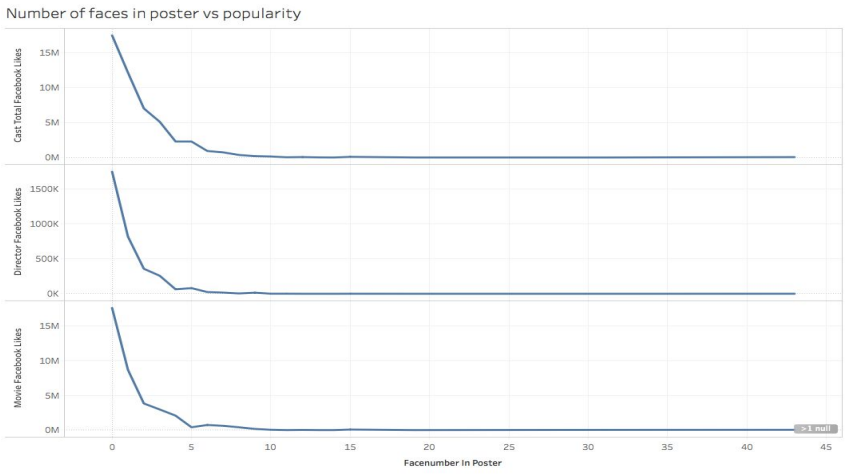

Comparison between IMDB score , Facebook Likes and Number of Awards per movie
Movie Title
12 Years a Slave
A Beautiful Mind
A Man for All Seasons
A Passage to India
A Streetcar Named ..
Amadeus
American Beauty
Annie Hall
As Good as It Gets
Black Swan
Blue Jasmine
Boyhood
Boys Don't Cry
Bridge of Spies
Capote
Charly
Chicago
Coal Miner's Daught..
Cold Mountain
Crazy Heart
Dallas Buyers Club
Dead Man Walking
Django Unchained
Dreamgirls
Driving Miss Daisy
Ed Wood
Elmer Gantry
Erin Brockovich
Forrest Gump
From Here to Eternity
Winner
Avg. Imdb Score
Avg. Movie Facebook Likes
Title Year
1934
2015
Movie Title
(All)
Highlight Movie Title
Film


Comparison between Facebook Likes and Number of Awards per movie over years
Winner
Movie Facebook Likes
Title Year
Movie Title
Highlight Movie Title


Academy Award Winners
Name
Count of Awards
Fb Likes
Imdb Score
Average
SUM(Imdb Score)
6.400
9.200
SUM(Fb Likes)
189
49,000
Name
(All)
Anthony Hopkins
Celeste Holm
Charlize Theron
Award
Actor
Actor in a Leading R...
Actor in a Supporti...
Year
1934
2015
Actor
Actor in a Leading ..
Actor in a Support..
Actress

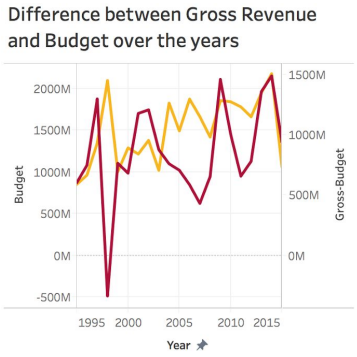

Academy Award Nominees
Name
Count of Nominations
Fb Likes
Imdb Score
Average
Meryl Streep
Kate Winslet
Denzel Washington
Leonardo DiCaprio
Robert De Niro
Anthony Hopkins
Jennifer Lawrence
Matt Damon
Morgan Freeman
Roy Scheider
Tom Hanks
Al Pacino
Albert Finney
Christian Bale
Claude Rains
Clint Eastwood
Ellen Burstyn
Holly Hunter
Jeff Bridges
Jeremy Renner
Johnny Depp
Laurence Olivier
Montgomery Clift
Naomi Watts
Philip Seymour Hoffman
Robert Duvall
Robin Williams
Sylvester Stallone
Tom Cruise
Tom Wilkinson
Will Smith
Amy Irving
Anna Kendrick
Anne Hathaway
Armin Mueller-Stahl
Benedict Cumberbatch
Bessie Love
Björnstjerne Bejo
Bill Murray
Brad Pitt
Bradley Cooper
CNT(Nominations)
SUM(Imdb Score)
6.000
15.600
SUM(Fb Likes)
28
49,000
Name
(All)
Al Pacino
Albert Finney
Award
(All)
Actor
Actor in a Leading R...
Year
1939
2015
Award
Actor
Actor in a Leading ..
Actor in a Support..

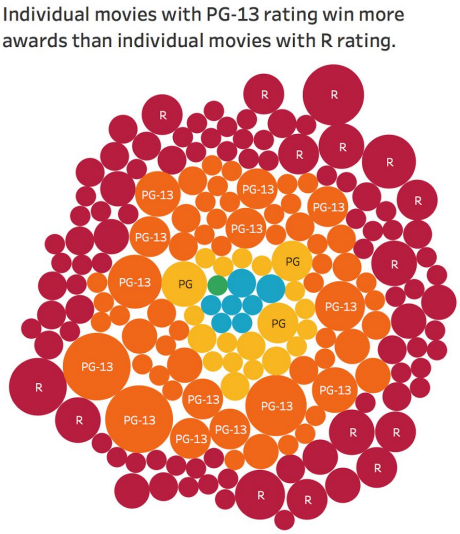

Budget and Gross vs IMDB score
Budget
Gross
Imdb Score
3 nulls
Measure Names
Budget
Gross

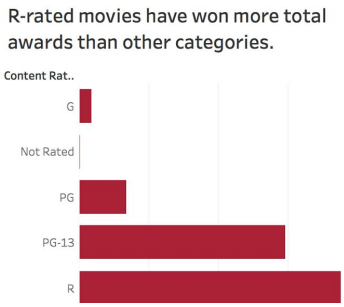

Number of faces in poster vs popularity
Cast Total Facebook Likes
Director Facebook Likes
Movie Facebook Likes
Facenumber in Poster
>1 null

There were too many factors and too many colors which makes these visualizations to throw off the users. Also, these designs did not capture what the users of our survey wanted to see, so we used these images to understand the data and scope it down and decided not to use these visualizations in our final design. We chose to use the data between the years 1995-2016.

We made another tableau visualization as shown below for our presentation but did not receive good feedback on it and also bubble size is not a good way to compare the numbers. So we decided to remove it as well.







We then started with the design of the visualization on paper first and brainstormed on different design ideas and different combination of columns in the dataset.

Data cleaning: We found that a lot of the data about actors in the movie were random and did not portray the top actors who would have won awards and so we ended up cleaning the dataset meticulously by scraping data from the IMDB website and replacing the erred data. We combined tables based on wins and nominations by actors, films, directors and year. This involved a lot of work.

Paper prototypes- We came up with paper prototypes before coding the actual visualization based on our ideas and tested it on users and got their feedback which enabled us to change the design and iterate through the design process efficiently. The paper prototypes are shown in the section above and shows how our design evolved over time.

We then conducted the survey to know more about the user preferences.

Summary of the key findings from the survey responses:

- **Narrowing down the focus**:

  What to focus on: From the survey responses, we could see patterns in what factors people take into consideration before watching a movie like IMDB score, genre, actor/actress, director etc and what are they really interested in knowing more about a particular movie/actor/actress/director like number of awards won and nominated for, director-actor pairing, average movie rating etc

  What to not focus on: From the responses that we got to our survey, we could conclude that some columns like facebook likes for movies, directors, actors etc were not what the people were interested in as all the survey responses unanimously showed this.

  Understanding the user base: From the survey responses, we could see that about 70% of the people watched movies as often as once a week to once every two weeks be it at home or in the theatre. This gave us a sense that movies is a topic that people can relate to quite easily and also proves huge popular interest in this domain.

We started coding in D3 and improved the designs further as per how the coding shaped up. We then integrated all the visualizations together on the same web page and added more contextual information around the visualizations and some interesting ‘Did you know?’ sections stating facts about movies and awards. The whole purpose was to create a coherent webpage that tied together all the visualizations and gave a good overall experience to the user and help him/her convey the story appropriately.

We then conducted usability testing on our visualizations and incorporated all the feedback that we got to improve the design further. The usability tests helped us identify the gaps in our designs and also the strengths of our designs. We incorporated the feedback in our designs and improved upon them.

7. Usability tests and responses

Usability test questions

Initial task: Please go through the webpage to get a feel of what it has to offer. Please let us know your comments/questions if you have any before we begin with the tasks.

Tasks related to visualizations:
Visualization 1:

1. What is the IMDB score of the movie Titanic?
2. For an actor of your choice, find out-
    a. the number of movies he has acted in
    b. his movie with the highest budget
    c. his movie with the highest number of awards won
3. For the genre drama, find out the movie with the highest budget and what is the IMDB score of this movie?

Visualization 2:

1. Which is the genre for which maximum number of actors have won awards?
2. For an actor of your choice, which are the genres that he has won awards for?
3. For a genre of your choice how many actors have won awards?

Visualization 3:

1. Name a director for whom multiple actors have won awards.
2. Can you locate where the directors are placed and where the actors are placed?
3. For an actor of your choice, who was the corresponding director?
4. For a director of your choice who is/are the actor(s)/actresses(s) who have won awards?

Responses

Usability test with user 1

Description of the user

This user is a Masters student and is a movie buff- watches all kinds of Hollywood movies in the theatre and at home. The frequency though changes as per her schedule of classes and work, she makes sure to catch up with all the latest movies whenever she has the time. She also watches older movies on Netflix whenever she wants to unwind.

Usability test responses

This user liked the over feel of the webpage including the color palette, font colors and types, small blurbs of text as her first response was that the color palette matches a lot with the Oscar ceremony. She related the “red” color used in the webpage with the red carpet and also noticed the gold colored Oscar award in the background.

For the first visualization she was able to do all of the tasks in very little time. She asked a few follow up questions around why the year of release for the movie doesn’t show up in the tooltip along with all the other information upon mouse hover on the circles in the graph. According to her, all the information included in the tooltip was important and said having the year there would have just completed it.

She could clearly differentiate between what the different sizes of the circles meant and what the different shades of the circles meant. She explored this visualization a lot entering different values in the search boxes and selecting different values in the dropdown.

For the second visualization, she was really impressed with the visualization. She liked the fact that we restricted the information to only actors and genres. She could complete both the tasks without any problems and liked that the two way relationship is shown. She could find the information correctly from the tooltips upon mouse hover.

For the third visualization, she again was happy with the interactivity of the visualization. She had a question though- “why are there just one director and one actor combinations?” but in another 3 seconds when she searched for Clint Eastwood she could see the 1 director to 4 actors relationship. She said that the small blurb of text that mentions this fact about Clint Eastwood actually helped her in expecting one to many relationships as well and she deliberately selected Clint Eastwood from the dropdown to check this. She took slightly more time (~ 20 seconds more) in completing the tasks only because of the different colors. She missed the key that showed the color assignment to director and actor as her feedback was that the visualization was huger and in the center as compared to the legend and so she missed it.

Usability test with user 2

Description of the user

This user watches movies that belong to mostly action and comedy genre. His movie watching habits are not very severe but claims to have watched all the award winning action and comedy Hollywood movies. He takes to Netflix to catch up with movies if he missed any.

Usability test responses

He went over the entire page scrolling top to bottom a few times before we was ready to do the tasks. He was reading all the information in text very sincerely. He liked the overall look and feel of the web page and liked the fact that we provided some contextual information before every visualization and this according to him was a good tactic to “separate one visualization from the other”.

For the first visualization, he was quite impressed with the zooming of the graph feature. He did take an extra 3 seconds to fathom what was on X axis and Y axis of the graph. He could perform all of the tasks quickly and correctly. He faced no problems in understanding the color legend, the size of the circles, the search boxes and the dropdowns.

For the second visualization, he could perform all the tasks as required. He also found the tooltip interesting especially when there were multiple movies within the tooltip. He was a little surprised at the number of the actors and actresses who have won the awards. But when I clarified that there are totally 82 actors/actresses and that we have considered only the 4 categories awards that concentrate solely on the actor’s/actress’ performances. This answer made a lot of sense for him and he liked the circular design idea as it could accommodate 82 names without taking a lot of space. Also, the ordering of the names in alphabetical order was very intuitive and helpful to him in spotting his favorite actor’s name.

For the third visualization, he could perform all the tasks correctly. He could spot the key of director and actor color assignment easily. He took lesser time in searching for the name of the director from the dropdown than from the visualization itself. He really liked this particular visualization as this was very interactive and all of the information that he wanted was “right there”. He also liked the fact that this concentrates only on the award winning actors and the corresponding director and not including award winning directors has kept this simpler.

Usability test with user 3

Description of user 3

This user doesn’t watch movies regularly but is certainly interested in movies. He divides his free time between movies, music, games etc.

Usability test responses

Upon just scrolling once through the entire webpage he got a good idea about what was going on. He liked the colors used and the interesting facts that were displayed before every visualization.

For the first visualization, he could perform all of the tasks in reasonable amount of time. He could easily search for movies/actors/directors easily in the drop down and also could select the said genres. He could easily differentiate between the budget and the number of awards won represented by the color and the size of the circles.

For the second visualization, he could again perform all the tasks as required. He could find the actors corresponding to a particular genre and the genres corresponding to a particular actor. He also found the tooltip helpful and could clearly understand the information displayed in the tooltip.

For the third visualization, he found the chart really interesting and was very impressed with it. He found the interaction seamless and found the color coding of the lines that join the director to the actor to be helpful. He too found the small blurb that we included above the visualization that mentions Clint Eastwood to be helpful and this told him that even though there are many one to one relations, there could be one to many too and this intrigued him to actually find those out. He went through all of the director names to find such pairings.

8. Links to demos, documents, or whatever is needed to show the visualization

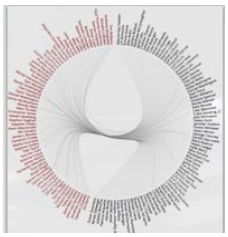
LINK TO PROJECT WEBPAGE

9. A table showing which parts were done by each student, and in what proportion approximately

	Astika	**Priyanka**	**Nia**
Obtaining data, web scraping, Pre-processing and clean up	20%	60%	20%
EDA using Tableau	33.33%	33.33%	33.33%
Initial Survey	33.33%	33.33%	33.33%
Initial Design/Design ideas	33.33%	33.33%	33.33%

discussion/Final idea decisions			
Visualization 1 coding	10%	80%	10%
Visualization 2 coding	80%	-	20%
Visualization 3 coding	20%	-	80%
Overall website design	15%	35%	50%
Usability studies	60%	20%	20%
Report	33.33%	33.33%	33.33%

10. A small thumbnail image (100x100 pixels)



11. Software

LINK TO CODE

13. Video of the visualization in action

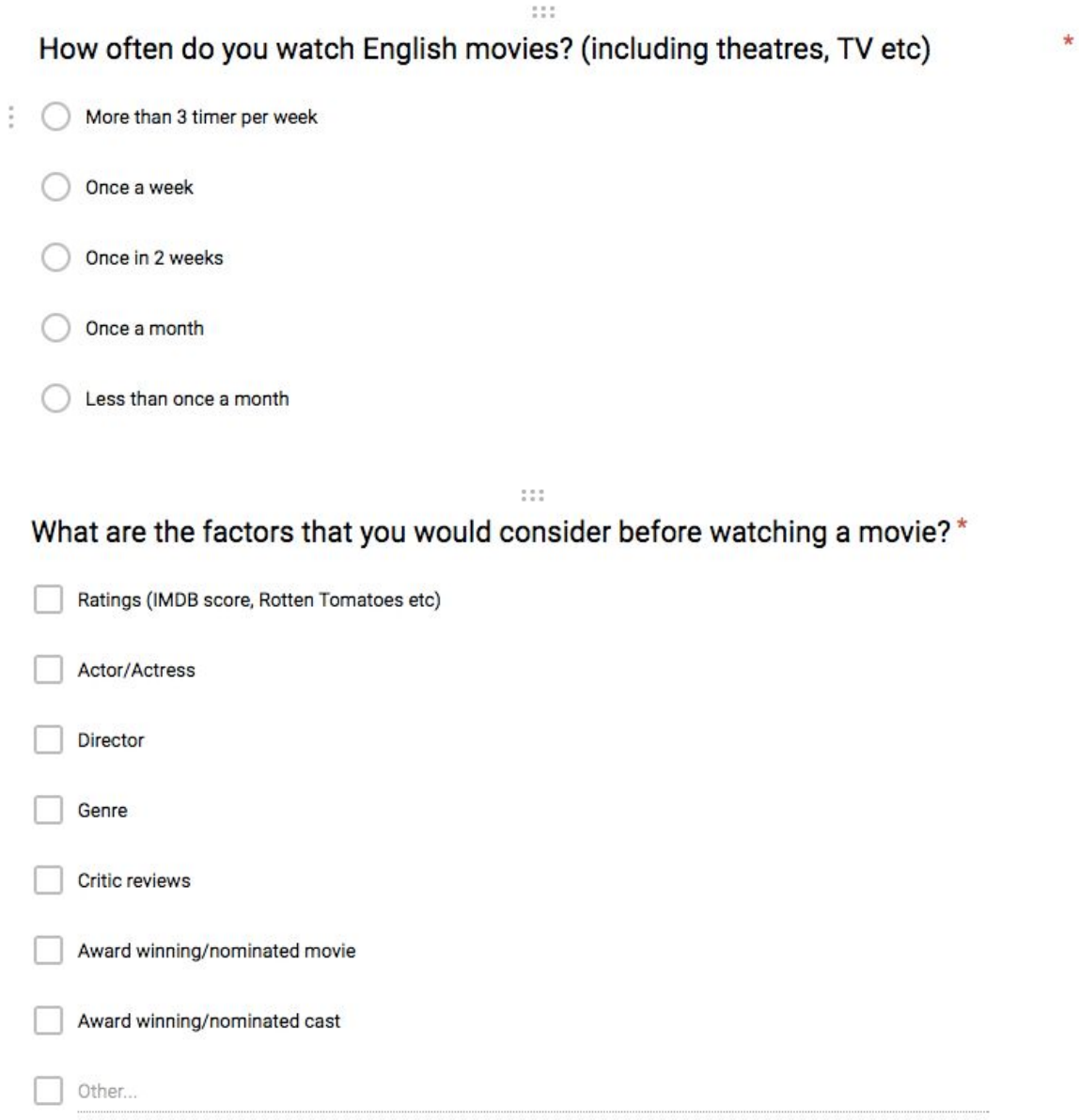
VIDEO OF DEMO

14. Appendix

- Survey-

  Link- https://goo.gl/forms/ChHR5JwheJbq1cSo2

  Screenshots-



What do you want to know more about an actor/actress? *

- [ ] Number of awards won/nominated for
- [ ] Genre most acted in
- [ ] Pairing with other actors
- [ ] Pairing with directors
- [ ] Average movie rating
- [ ] Movies acted
- [ ] Facebook popularity
- [ ] Other...

What do you think would increase the chances of a movie winning an academy award? *

- [ ] Actor/actress
- [ ] Director
- [ ] Genre
- [ ] Popularity of the movie
- [ ] Budget
- [ ] Other...

What would you want to know more about? *

1. Movies
2. Actor/Actress
3. Director
4. Genre

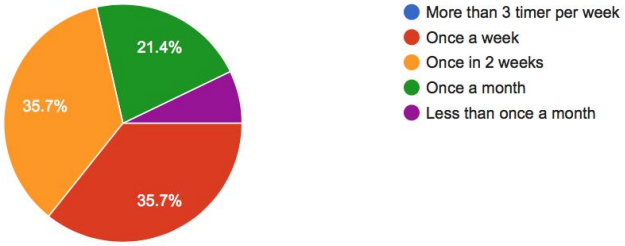
Is there anything else that you would like to have in a movie visualization that we didn't cover above?

Long-answer text

RESPONSES FOR INITIAL SURVEY

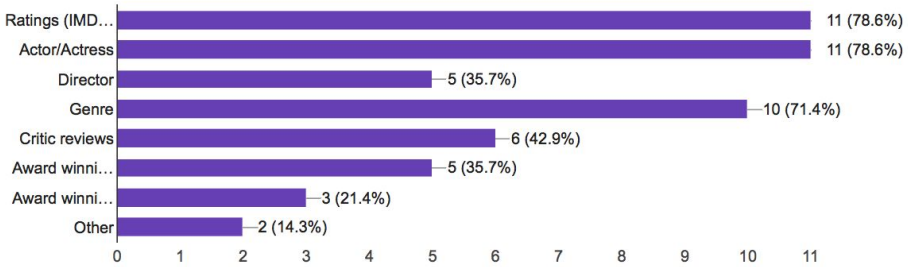
How often do you watch English movies? (including theatres, TV etc)

14 responses



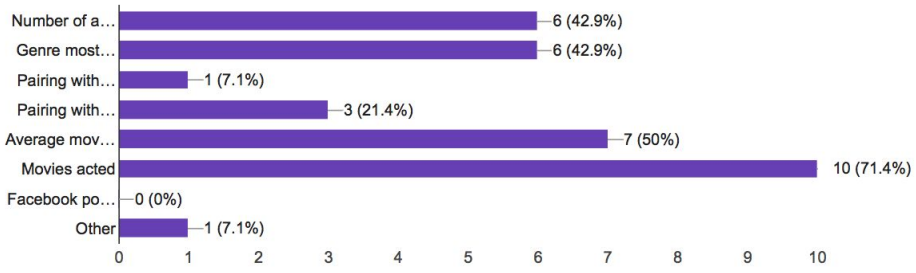
What are the factors that you would consider before watching a movie?

14 responses



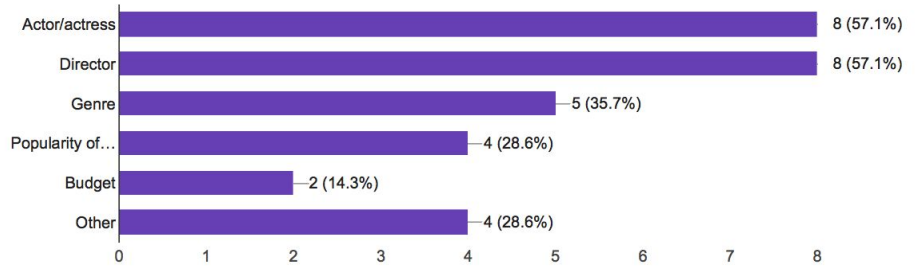
What do you want to know more about an actor/actress?

14 responses



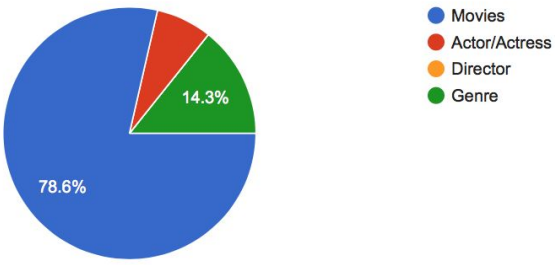
What do you think would increase the chances of a movie winning an academy award?

14 responses



What would you want to know more about?

14 responses



Is there anything else that you would like to have in a movie visualization that we didn't cover above?

6 responses

plot tropes
maybe first week /fourth week delta (if something started low and spread through word of mouth to make more money later, i'd want to know about it)
movie suggestions like pairs or groups of movies that a person would be interested in based on a particular movie that the individual likes
If the movie is based on a very famous novel then is that popular?
The effects or technology used in a movie e.g. 3d vs 2d, also affects my decision to watch or not to watch a movie
Movie length, Producer