SI649w14 – Assignment 3 Final Report

Team member names: Chin-Jui Chen, Bingran Guo, Dinghua Liang, Hsiao-chih Lin

Project name: Vistocks

What to turn in

Answer **all** the following questions and provide screenshots in the document (so I can print out one file and you can explain what you're talking about, .pdf or word would be great).   In addition turn in your working code with instructions on how to run it (you're welcome to host the site wherever you want if you want it online, but please turn in .zip or .tgz file with the source). Note that the format of this has changed a little bit.

Problem definition, Phase 1

1. What is your user getting "paid" to do? (note that this doesn't need to be financial gain)

    > They are being paid to identify whether their virtual money should be spent on a specific MovieStock on HSX.com.

    a. Who are they?
    > HSX investors who have prospective MovieStocks in mind and especially for those who are interested in investing the recent IPO MovieStocks.
    b. What are their needs?
    > Long-term needs:
    > Have fun and make good investment, earn virtual money in H$, and have a sense of accomplishment.
    >
    > Short-term needs:
    > Evaluate the value of a specific MovieStock
    > See the distribution of casts' expertise and relevant experiences in the movie industry. Find the cast member(s) who have greater impact both in good and bad over the cast and the movie.

    c. What are their priorities?
    > Their goals is to evaluate a MovieStock and decide whether to invest it or not, and the dimensions they evaluate are vary from person to person. However, overall their priorities are evaluating the perspectives which benefit a specific MovieStock and identifying specific facts that will increase their belief on a specific MovieStock value in the future.
    > In a more objective view of evaluating a MovieStock, users need to see figures of

MovieStock's current values, ratings of the cast members, box office, and etc. To be more specific, users will need specific strengths of each director/actor, specific number of works in different genres and years.

d. What are their limitations?

They may not familiar with some of the actors or directors. They are also usually more familiar with the images and the movies an actor performed before. We need to provide cues to help them link back to the things they are familiar with.

They may new to HSX and making their first investment. In this case, users have no preference on which dimensions they evaluate more than the other for investment. Besides, each MovieStock has different influential factors that drive the changes of the stock value, we should help users to find out the key factors easily.

e. What are their biases?

Each person has his/her own taste of different types of movies and actors. People usually don't judge a person very objectively. We will need a neutral, objective, and credible source data to support our end-users to evaluate a MovieStock.

Before seeing the cast visualization, they might assume some of the experienced actor/actresses would have higher market value than the some young stars, or the other way around. For example, Emma Watson's Google trends popularity is way higher than Anthony Hopkins, but the current value of Anthony Hopkins is higher than her.

f. What do you *not* know about them?

We do not know whether the investors invest based on more of personal preference or objective informations, such as current values, ratings from an authoritative movie, and etc... They may need to discuss with more people, either professional or amateur, before they make an investment decision. We don't know how seriously they treat the investment on the Hollywood Stock Exchange. They may not care about the outcome and make decisions without a thought.

2. What questions do they need answer or decisions do they need to make as part of their "job?" Think of these as the main subtasks. These should still be "domain" specific.

In my role as a MovieStock investor I need to identify whether my virtual money should be spent on this MovieStock:

a. Is the combination of director and cast benefit to this movie or not from the perspective of objective statistics?

    i. What are the potential values of each cast member of this movie? (The stock price(H$) of cast member)

    ii. Do these cast members have expertise in the relevant genre of this movie, or vice versa? (Ratings, number of works, and past works in relevant genres)

    iii. Do these cast members have cowork experience with each other and how tight is that?

b. Do I have personal preference of these cast members based on my past experience? (from their profile pictures, names, and their past works)
c. How popular are the movie itself and the cast members over the past few months on Google Trends and how will that affects the MovieStock price?

3. What "statistical" or "analytical" operations are required to achieve the tasks in (2)? These are things like comparisons, trend identification, outliers, etc. At this point we're moving into the domain-independent realm. Things like comparison, ranking, outlier detection, etc. This is also where you define the mapping between the domain specific (e.g., flights) to something more generic (e.g., time series analysis).
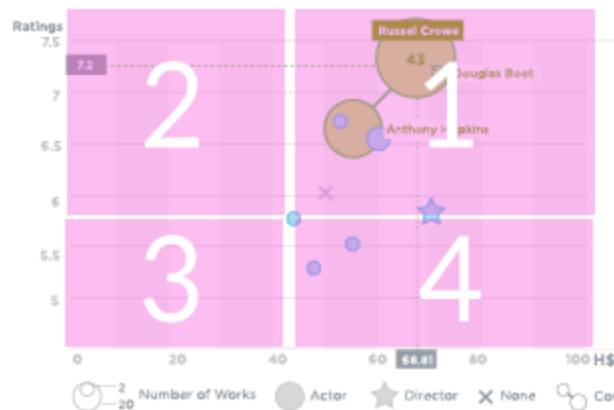

Fig 2.

To answer question a, users identify a trend of scatter plot to see where do the cast members fall on to. For example, the cast members in Area 1 would be perceived higher potential to invest since they are both highly priced and highly rated. The cast members in Area 2 or 4 are less attractive but still valuable because they are either highly priced or highly rated. If cast members mostly locate in the Area 3, the movie would have the least potential to be invested.

Users utilize the same analytical operations to see if the scatter plot of the relevant genre indicates considerable potential to be invested. Users can perform a outlier detection, to identify whether there are actors who are dominating this movie cast, or who have no experience in the relevant genre. If the cast members have less or no experience, the investment of this MovieStock is relatively risky.

Pairs of actors or directors who have co-worked experience before will be evaluate more important and provides some faith on the future performance of this movie.

To answer question b, we need to add a way for users to quickly identify the cast members and link back to their memory. Actors or directors who are more familiar to the users will be weighted more important than the others.

To answer question c, mapping the history of hotness of movie, director, and actors with the stock price history line chart helps users to judge the correlation between popularity and the stock price, and also to estimate future price trend.

Implementation, Phase 2

1. What data did you end up using?

    List of trading MovieStocks from HSX.com
    List of cast members (actors, actresses, and director) and their HSX prices from HSX.com
    Box Office of each cast member (in HSX cast members) from Boxofficemojo.com
    List of past works of each cast member (in HSX cast members) from IMDB.com
    Movie genre and rating of each past work from IMDB.com

Categorical	**Movie genre**	HSX
	Cast members (Actor, actresses, and Director)	HSX
	Rising or dropping value of a cast member	HSX
Nominal	**Past works of each cast member**	IMDB
Ordinal	**Distribution time of each movie**	HSX
	Time series from Google Trends	Google
Quantitative	**HSX MovieStocks value (H$)**	HSX
	Number of works in total/genre of each cast member	IMDB
	Average Box Office of past works of each cast member	Boxofficemojo
	Average Ratings of past works by genres of each cast member	IMDB
	Index of Search Interest from Google Trends	Google
Relational	**Co-worker Experience between two cast members***	HSX/IMDB

2. What did you do? (**provide screenshots to explain what you're talking about**)

Fig. 1. Visual Encodings

○ Encoding

There are two parts of the visualization: the upper part is the cast visualization, and the bottom is the Google Trends.

In the cast visualization, there are the scattered plot and the side panel.

In the scattered plot, the y-pos encodes the average rating. The x-pos encodes the stock price. The color green indicates where the user highlights, and which two agents have worked together in the past. The size of the circle indicates the number of works in total or relevant genre, whereas the shape indicates the type of cast or the type of data. The enclosured glyph of circle and star each represents actors and directors. The cross glyph, fundamentally different from the other two glyphs, indicates that the actor has zero number of works.

In the side panel, the color and the orientation of the arrow indicates whether the price of an actor is rising or dropping. In the bar chart, y-position encodes the actor's expertise from top to bottom, the x-position encodes the rating in each genre. In the list of past works, the y-position encodes the recency of works from top to bottom.

In the Google Trends visualization, the color encodes each cast members, x-position encodes the search interest, and the x-position encodes the time series.

○ Interaction [1]
1. **Connect**: Reveal data or relation that's originally not shown
   Hover:
   Highlighting a circle also highlights the profile picture of the corresponding cast member. According to Gestalt principle, Common Fate, the elements change together are easily perceived as the same unit.This interaction helps users to link the concepts between the quantitative data from the scatter plot and the actor's profile picture.
   The principle also applies to the detailed information, like the number of price and

the number of ratings, which are associated with the selected cast member. Hover over also reveals the co-working experience from the past between two agents.

2.   **Select:** Retain an actor/actress that's interesting to the user
      Click: Once clicked, the detail panel is presented until the user select another agent either from the scattered plot or the side panel.

3.   **Filter**: Show data conditionally
      The tab on the top-right allows user to switch relevant genre to see the data.

4.   Exploration: In our demo, the webpage supports a dropdown list of the trading MovieStocks with pagination. This is not in the scope of our visualization but it allows users to find a MovieStock.

3. How do the implementation choices support the operations in Phase 1, part 3?
    Our implementation supports the trends identification because 1) the use of scatter plot, 2) the color highlights and the linkage indicating the co-work experience. To elaborate that, the size of the circle draws attention to the cast members who are more experienced. The nature of the scatter plot enables users to identify where the cast is locate, for example, are they aggregated in the highly rated and highly priced area? The color highlights bring users attentions to the actors who have worked together in the past. This implementation helps users to see the relationship and gain insights on whether the MovieStock is valuable.

    To support the recall of the past experiences of actors and movies, we add a side panel showing all the pictures of cast members. Also, past works are shown in the detail page of a cast member.

    To support the trend identification of Search Interest, the line chart from Google Trends illustrates the popularity across the history, so users can have an idea of how the popularity of the movie or cast members changes over time. This implementation allows users to predict the value of the MovieStock from observe the overall trends.

4. What doesn't work?
    ○   The design layout doesn't support a list of cast member that is more than ten. This is a minority in the dataset as most of cast list is between one to five members.
    ○   The correlation highlights overlap among each other and are difficult to distinguish when circles are crowded in an close area.
    ○   History data can be used to analysis the correlation of variables and HSX price to make our visualization more credible. However, the nature of the history is scattered because active years of each actor/director are different and cannot be normalized on the same graph.

    Reflect honestly (but briefly) on what didn't work or you didn't finish.

1. How did you evaluate your solution? If you didn't evaluate, describe an evaluation plan.

    We evaluated the early visualization by cognitive walkthrough with the peers in our class. Our team gave a pitch of the background and use case, and asked for their feedbacks. We have also conducted internal critique to keep the design rationale reflecting a neutrality and flexibility to serve two types of personas: the enthusiasts and the analysts.
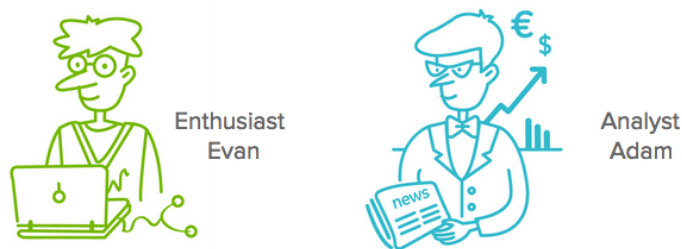

    Fig.3 Two types of Personas [2]

2. If you did an evaluation, what did you find?

    We have gained insights from our peers. First, the clarity of the language used is improved. For example, for "H$" we added the label to clarify it indicates the stock price. Second, we incorporate a visualization of expertise with horizontal bar chart in the side panel. It was one of our early ideas, but discarded because we considered the information might be trivial to the users. However, in one round of our peer feedbacks, they brought up that they would want to see the expertise distribution of a particular actor/actress. We hence brought the feature back to the final version. Third, we incorporate the feature of highlighting the co-working experience between two actors. Same with the expertise bar chart mentioned previously, we had the idea earlier, and the feedbacks helped us to bring back the feature.

Additional reflections

1. Any other thoughts about what worked or didn't work? What you would have done differently?

    ○ The stacked area chart of search interest of the cast members does not work because there were difficulty retrieving the data from Google Trends.
    ○ Real-time updates with HSX.com and IMDB.
    ○ Cross domain problem would arise when our project is not on HSX.com domain, and our demo during presentation works fine because we run it on our local machine, while when we put everything in an actual server, like the one offered by people.si.umich, you would see the cross domain issue.
    ○ We also thought of adding search bar to access the movie. This might be helpful when the user want to quickly navigation to a specific movie.

2. Were there alternatives you considered that you ended up tossing? (feel free to include screenshots)

An alternative we have considered presenting is incorporating a url "viz-hsx.com". Apart from the dropdown list of MovieStocks we use in the final version, users locate a MovieStock on HSX.com and add the abbreviation in front of the url to brings up a page with our add-on. After discussion, our team thought it requires an additional cost for users to type in the url. Thus we went with the demo with simple navigation to go through a list of MovieStocks.

Reference

[1] J.S. Yi, Y.A. Kang, J.T. Stasko and J.A. Jacko, "Toward a Deeper Understanding of the Role of Interaction in Information Visualization", IEEE Transactions on Visualization and Computer Graphics, Vol. 13, No. 6, Nov/Dec 2007, pp. 1224-1231.

[2] Illustrations retrieved from http://newdami.com

Appendix

[1] Presentation slides

[2] Demo Site