# CHRIS LEUER, CSCI-171 PROCESS BOOK
BOX OFFICE SUCCESS VISUALIZATION, APRIL 2014

INSPIRATION

Hans Rosling's Wealth & Health of Nations presentation is the original inspiration for the use of a time-series visualization to show box office success. I was interested in creating animation that shows a trend over time. Below is a re-creation of this visualization in d3.js Mike Bostock. *http://bost.ocks.org/mike/nations/*
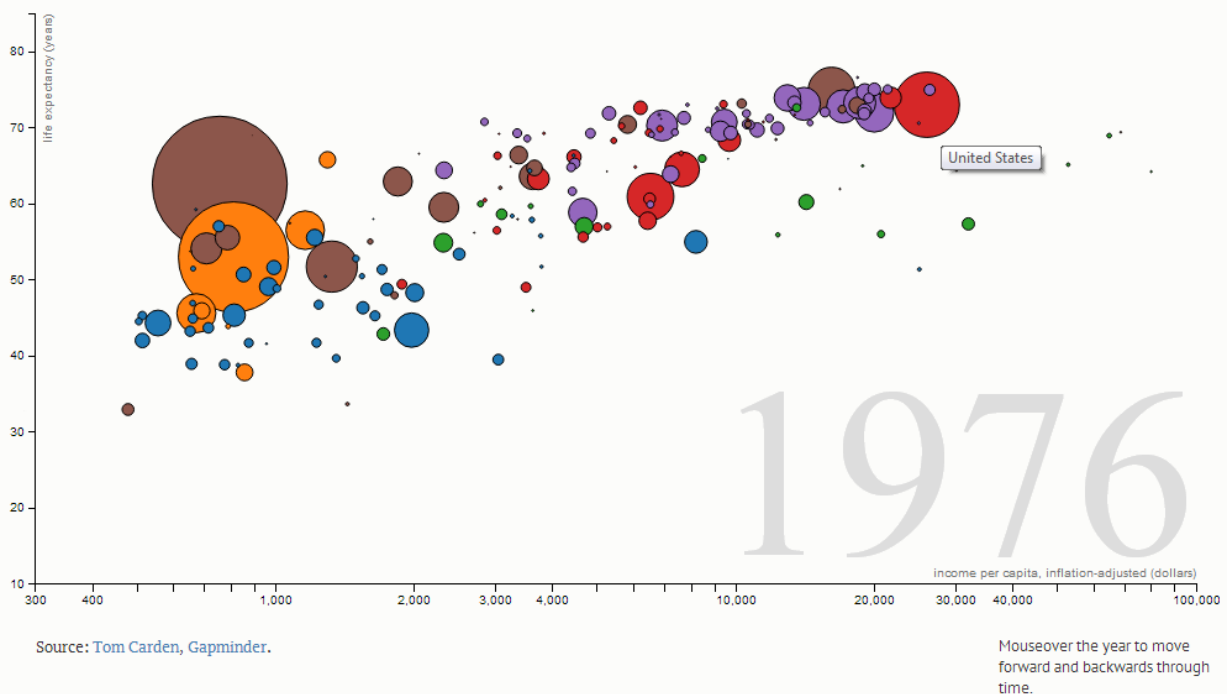

I dreamed of creating my own time-series animation on the fun topic of movies. I wanted some level of interaction to explore different dimensions of movie success.

UNDERSTANDING THE DIMENSIONS

Position is one the strongest visual patterns for differentiation.   Here, the height of movie represents the box office success as a measure of weekly gross. Since this is animation, the position will change over time. A mega-successful movie like Hunger Games or Skyfall will have a high starting position in week one far above the rest of the pack and will stay above many other movies , even those in week zero. This allows the viewer to quickly identify success movies. Hunger Games is still making over $20 million in the fourth week.

Furthermore, the viewer can stop animation to investigate in more detail why the movie is successful.   By default, the animation draws the circle radius as function of theatre counts.

Theatre count is another measure of success which shows staying power.   A movie earns a larger gross if it stays popular in movies for a longer period. The user can then has the option to compare different factors which may contribute to it's success. Budget is a good sign a movie will have a high theatre circulation. Facebook likes is an indication of social media activity.   Rotten Tomatoes critics score gives an idea of how good the reviews were on TV and other media.   Rotten Tomatoes audience is an indication of how well like a movie is by the general audience.
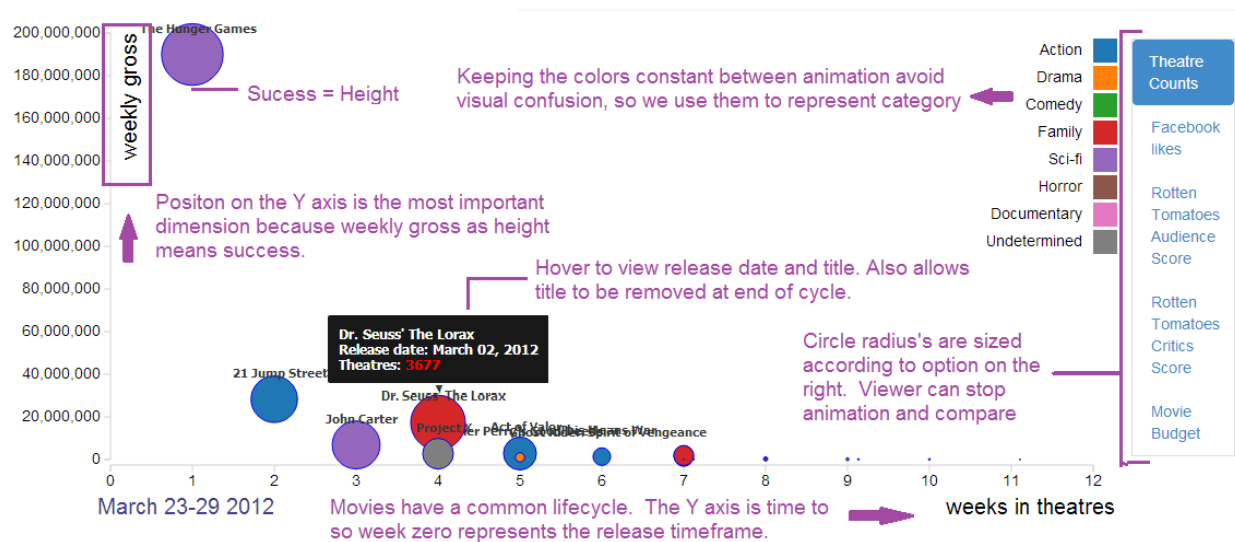
EVALUATION

- Evaluation: What did you learn about the data by using your visualizations? How did you answer your questions? How well does your visualization work, and how could you further improve it?

DATA

I learned Facebook likes as a representation of social media activity range across a big spectrum and do not have a strong correlation to movie success. Generally, family animated movies have large following on Facebook. These movie do predictably well, but there large difference in the amount of likes for two blockbusters. Ted is an example of comedy that did remarkably well but does not have a large Facebook count.

The box office mojo data turned out be excellent and high quality despite the screen-scraping. The problem is I ran out of time to incorporate the 2013 results. The API from Rotten Tomatoes was excellent too. The problem was coming up with a master list since I couldn't easily get category. In the future years, I need to find a way to reliably get likes and category.
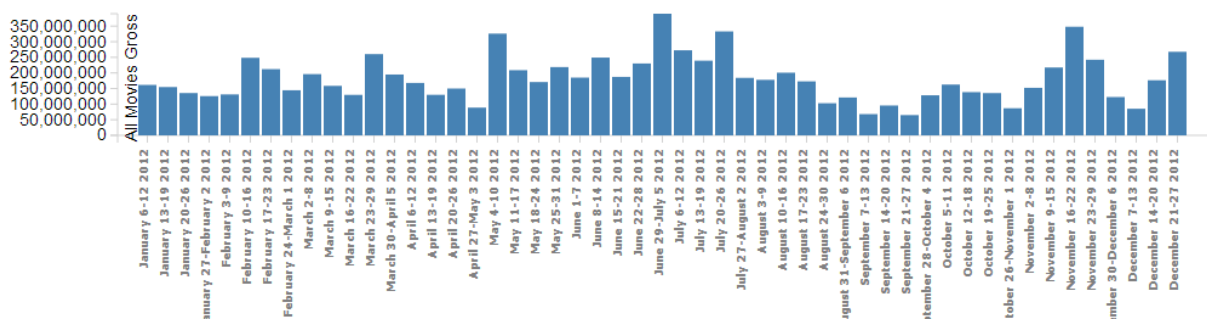
IMPROVEMENTS

I think this visualization would be more exploratory and fun to interact with if there was third chart to show Gross Total by movie. By clicking on a movie it would initiate the animation for this movie. Also, the animation

ANIMATION

Getting the animations right was a huge time investment. I didn't really have an example to work from which could layer data. It turns out this is very challenging. I am new to Javascript and d3.js so there surely are improvements to made to the code to make is more re-usable and less buggy. There are issues around stopping an starting animations.

RE-THINKING THE SECONDARY CHART

Originally, I drew a line chart as the secondary chart thinking I would use a brush to move the animation to different periods. However, this would have been a poor visualization since the number of movies crammed into a the narrow height would have made it unreadable. A much simpler implementation is actually better because you get an idea what periods had the biggest results. Below the summer blockbusters start to arrive in May. The results taper off in fall.
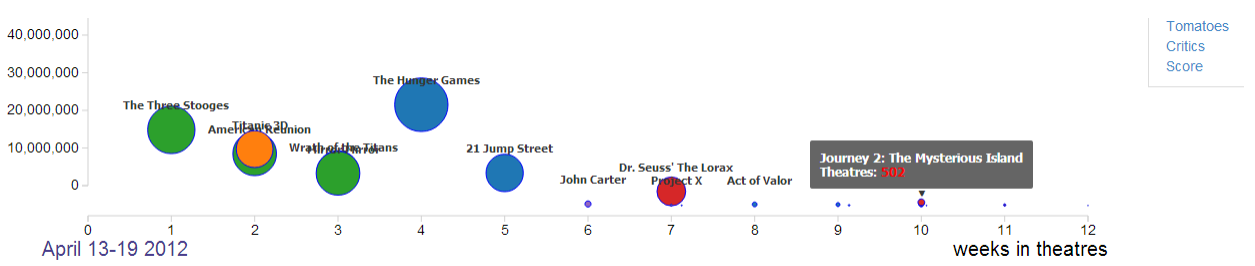
TITLE

Movie which stay in theatres past 6 weeks typically make less than 10 million. Therefore, they are no longer readable as they use the same Y position.   Title should be removed.
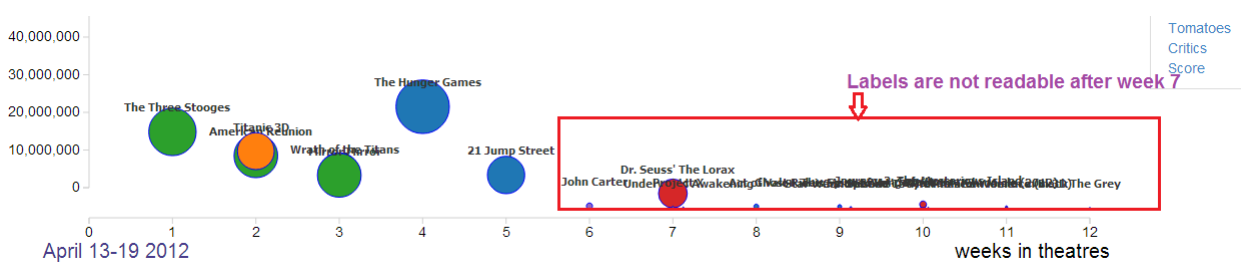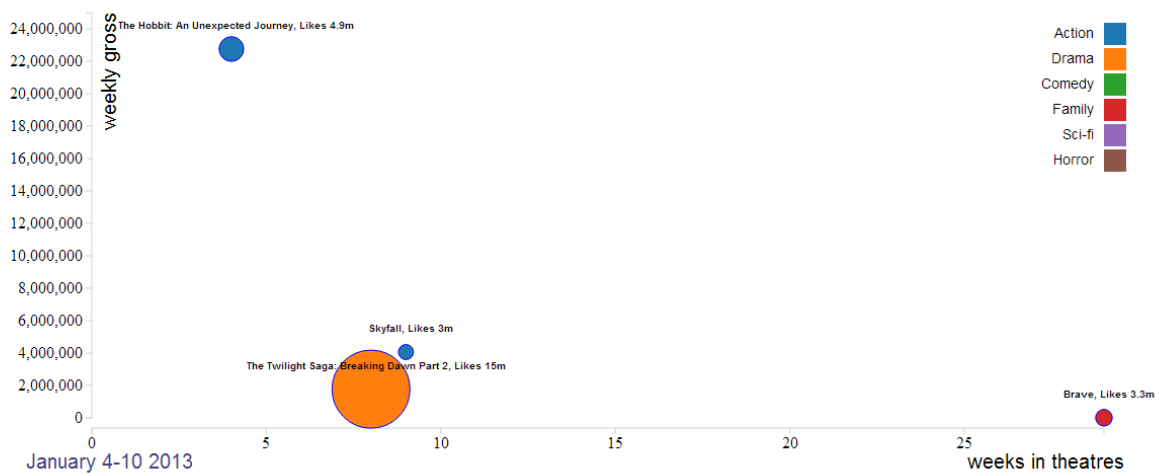

The improvement are mouse-over labels on circle to display title name, release date and quantity representing radius size.

PROTOTYPE

Box Office Success

This is static prototype of the main chart. When animated, the movie circle will move left to right, as the number of weeks increase.

INTERACTION

4/10 - It might be a useful if the user should be able to toggle between weekly gross and total gross, forcing a re-animation.

4/10 – Also, the url is in the data, so I plan to link to an external page in boxofficemojo.com. It shows a breakdown of statistics.

CATEGORY DIMENSION

4/9 - Since the final data may be very rich with many movies in the scatter plot, it's not visually feasible to color according to movie title. Instead, I plan to represent category as color.

Drama, Sci-fi, Comedy, Action, Famly, Horror

ANIMATION

THE MAIN SCATTER PLOT WILL AUTOMATICALLY TRANSITION FROM THROUGH 4 WEEKS OF BOX OFFICE RESULTS. IN EFFECT THERE WILL BE 4 STATES OF TRANSITION. **THIS WILL NOT BE FULLY IMPLEMENTED UNTIL AFTER PROTOTYPE.**

BRUSHING AND TOTALS BAR CHART

weekly Gross

Animated →▷ →➤

r = Facebook likes

4 Step Animation Hobbit

Hobbit

Hobbit

Hobbit

r

r

r =

Change brush forces re-animation

week 1   week 2   week 3   week 4   weeks

December 2013 ← brush?

2013

Filter by [movie category]

Brush

Hobbit   Hunger games

frozen

2012   2013
2013

total gross

Bar Chart by year?

2014

Hobbit
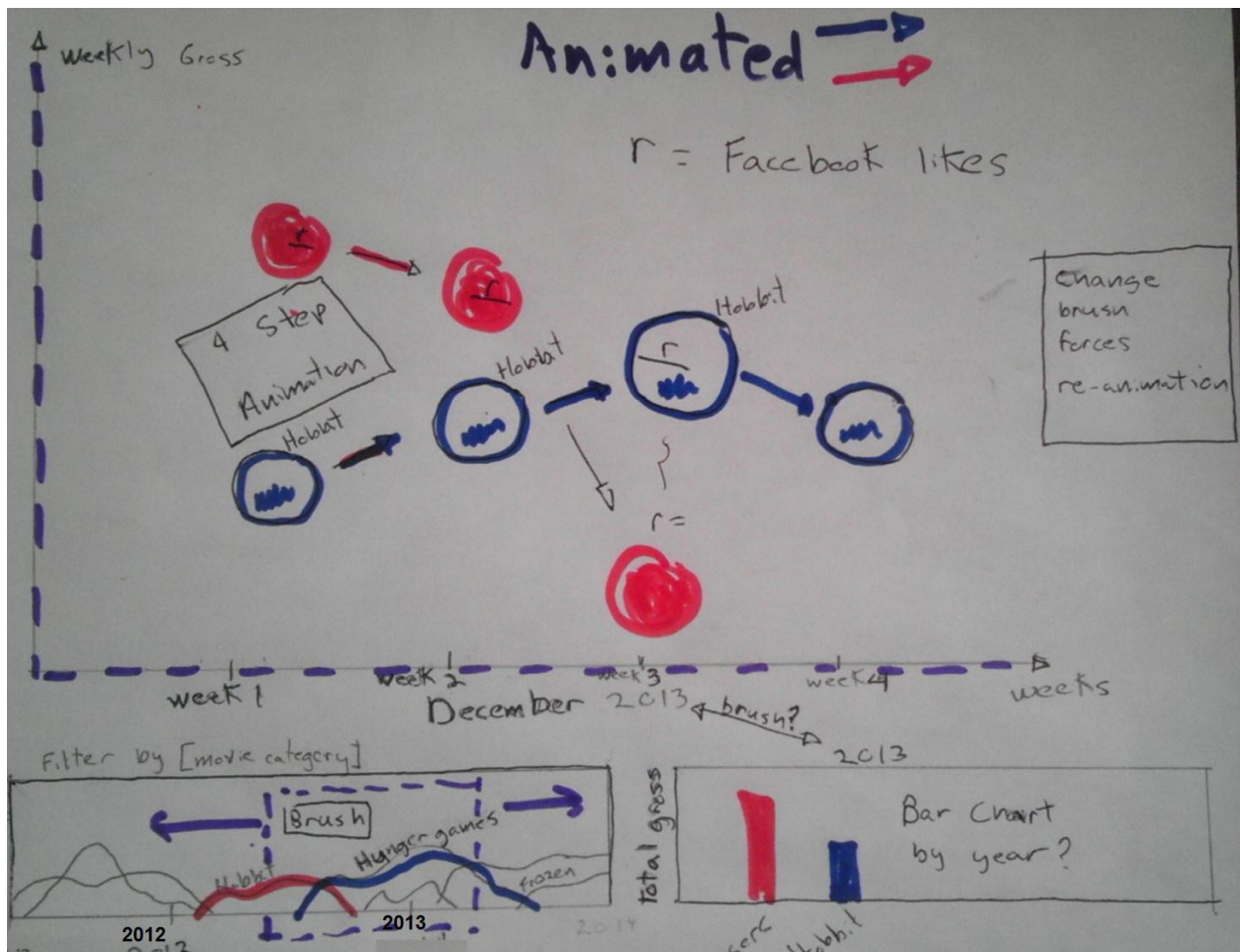
Here is the secondary line chart. Each line represents a movie and it's weekly gross along the Y axis.
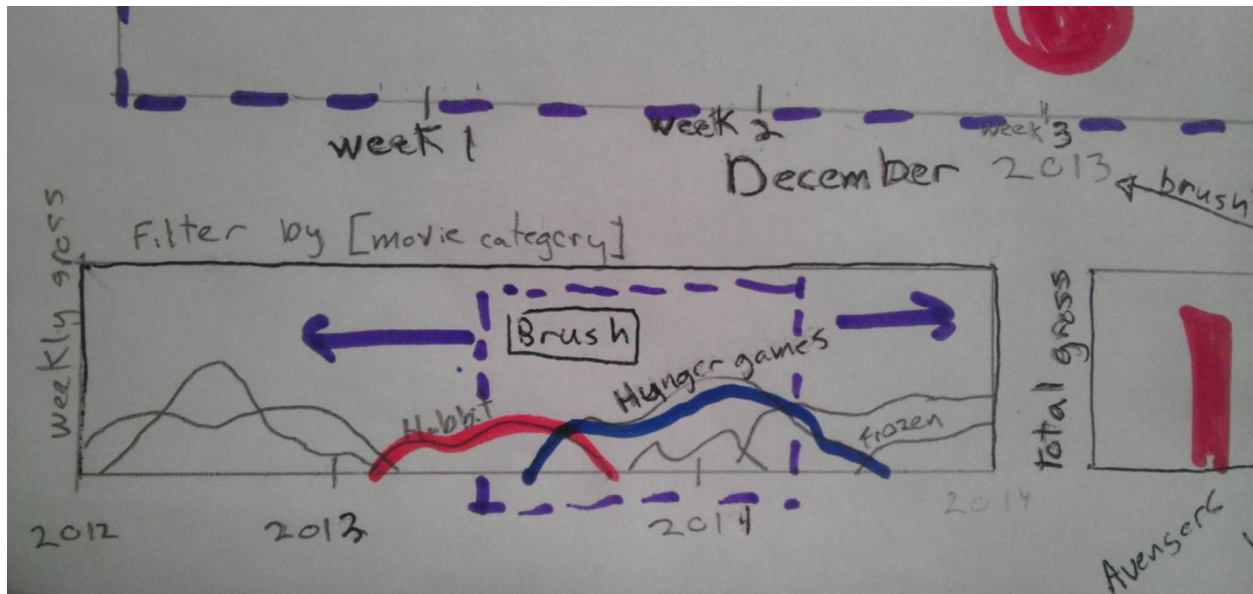
Social Media dimension

After playing around with it, I cannot find a way to use the twitter API does allow me to aggregate weekly mentions by a specified date. This would have allowed me to match social media activity to weekly gross for each movie. Instead I plan to use a simple social media metric, total Facebook's like for each movie. This will be measured by the count of likes for the official movie facebook page.

Facebook Likes can be visualized as the radius of movie circle.

DATA SET

movie-data array			
0	weekDates	weeklyGrossSum	**weekly array of records**

	weekDates	weeklyGrossSum
0	January 4-10, 2013	$ 300,000.00

weekly array of records

	title	weeklyGross	week	theatreCount	budget	likes	release date	likes
0	Django Unchained	$ 28,064,745.00	2	3,010	100m	10000	25-Dec-12	
1	Texas Chainsaw 3D	$ 25,601,740.00	1	2,654	-	12000	4-Jan-13	
2	The Hobbit: An Unexpected Journey	$ 22,754,680.00	4	3,755	-	30000	14-Dec-12	

PROPOSAL

CSCI-171 Data Visualization Project Proposal

Box Office Success and Social Media Activity

Chris Leuer, March 2014

- **Background and Motivation.** Discuss your motivations and reasons for choosing this project, especially any background or research interests that may have influenced your decision.

I am interested in working on a visualization which shows a universally interesting trend. Therefore, I propose a visualization to show movie box office success as it compares to social media activity over time.

I was inspired by Hans Rosling's 200 countries and 200 years. Since I am taking this course as point of personal curiosity, I want to work on a fun and universally understandable topic, and, well, I enjoy movies.

- **Project Objectives.** Provide the primary questions you are trying to answer with your visualization. What would you like to learn and accomplish? List the benefits.

The box off results visualization should show a positive correlation between social media activity and box office results. The more buzz a movie gets in Twitter the better it does. This is my hypothesis, but I am ready for the visualization to tell a different or more nuanced story. Of course, there are more factors involving box success, like budget and theatre circulation, and, of course, non-data factors like reviews/awards and movie content. As such I might also include budget, category and circulation as quantifiable dimensions.

Questions

Are there movies with a lot of social media buzz, which perform poorly relative to their high budgets? Will a movie with fall off dramatically after opening day as social media falls off? Are there movies with small budgets and high social media activity which perform well? Did some movies get no social activity but do relatively well?

Benefits

Mainly, I hope this visualization entertains.

This visualization might benefit movie executives who want to understand the role of social media in box office success. Should a studio invest heavily in non-traditional social media marketing campaigns? Will a lack of social activity doom a large-budget film.

- **Data.** From where and how are you collecting your data? If appropriate, provide a link to your data sources.

The data will come from two sources.

1. The source of social media activity data will be the Twitter API 1.1. The Twitter tweet Search API will be used to find movie mentions in weekly date ranges.

   https://dev.twitter.com/docs/api/1.1
   https://dev.twitter.com/docs/api/1.1/get/search/tweets

2. The source of Box Office results will use the Box Office Mojo website (unless I can find an API). Gross and other data like budget can be found in tabular form at the following site,

   http://www.boxofficemojo.com/

- **Data Processing.** Do you expect to do substantial data cleanup? What quantities do you plan to derive from your data? How will data processing be implemented?

    I do not expect any gaps in box office results or twitter search results, so I do not expect to estimate any missing data. Twitter tweets are 140 character messages, so my solution will need to perform a simple text pattern search to get number of tweets which mention a movie. I plan to limit the number of movies. I intend to use JQuery to read the various pages of movie results from HTML tables. The data is segregated by week into many pages so this will be challenging.

- **Visualization.** How will you display your data? Provide some general ideas that you have for the visualization design. Include sketches of your design.

    I am still thinking about the visualization, but the primary visualization might be a chart where each movie is circle.

    Primary dimensions I need to show are:

    1. Social media activity as number of tweet mentions in a week.
    2. Box office results, weekly and/or gross
    3. Timeframe in weekly magnitudes

    Secondary dimensions might be:

    1. Category of movie, i.e. science fiction, romantic comedy, drama
    2. Budget
    3. Number of theatre circulation
    4. Opening day

    This visualization might involve animation where the chart X-axis changes as function of weekly intervals. Categories can be represented as colors. Radius size is box office gross.
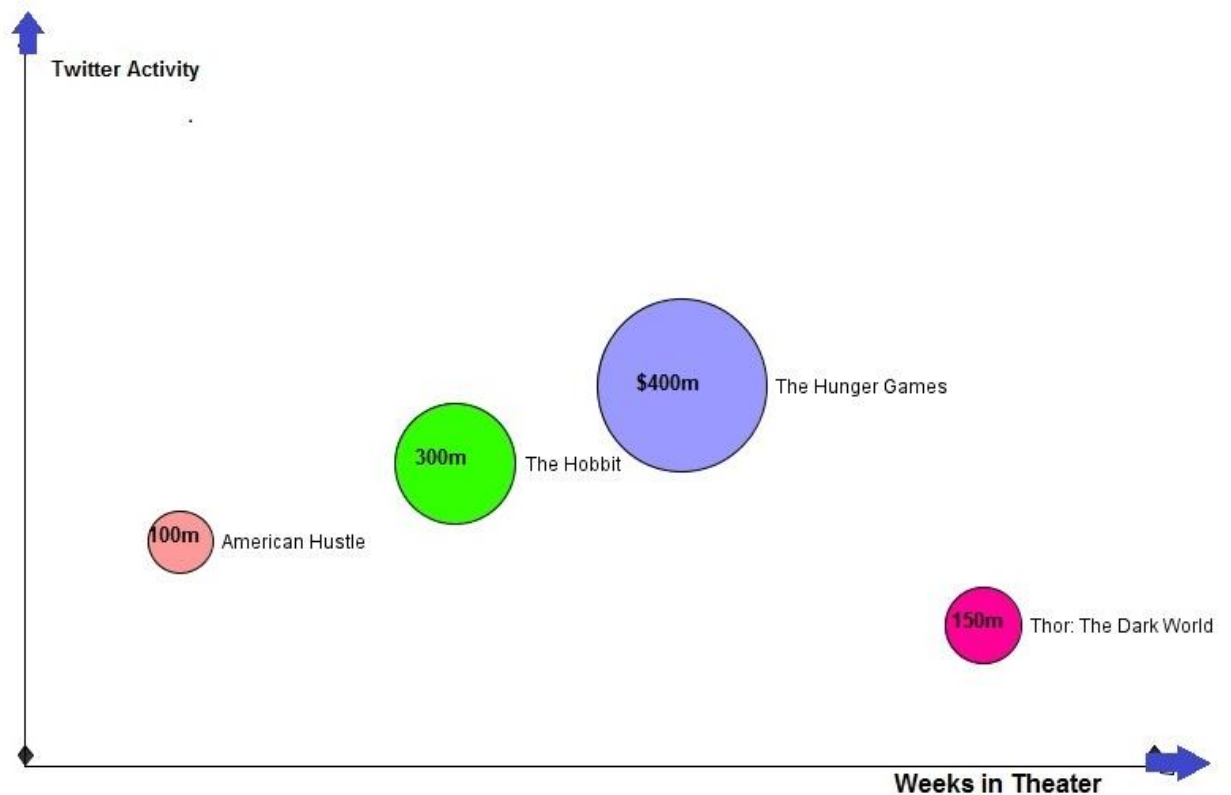
For another chart, or secondary chart, I also like the idea of a time-based flow diagram, where the colored segments represent the number of tweets relative to the total. This could potentially act as the overview (yearly time scale) for the more detailed weekly graph. This would require brushing.

- **Must-Have Features.** These are features without which you would consider your project to be a failure.

  The visualization must show the main dimensions or box office results, twitter activity and timeframe. The visualization also must show the full life cycle of enough movies to understand the trend of social media in movie success over time. There is no avoiding that some movies cycle will get cut short at the start and beginning of the dataset. A mega-hit can stay in theatres for 10 weeks while a poor performer will only last a couple weeks. Therefore, I probably need one year of data to show trends.

- **Optional Features.** Those features which you consider would be nice to have, but not critical.

  1. I would like to show movie budget and theatre circulation dimensions. I think these would be helpful in telling the large story of movie success. I movie which has $100 million budget and is in every theatre, but only earns $80 million is not a success. I definitely intend to show budget, even if just a hover over.

  2. Show movie events. I would like to show opening day.

  3. Some form of time based auto-animation would be really useful. Pick a year, watch the flow of social activity vs. movie gross.

  4. My dream scenario relates to #2. The user to click into an individual tweet as it animates (twinkles) for a few seconds as the timeline animates.

  (frankly , 1 and 2 are probably not feasible since I am one person)

Project Schedule. Make sure that you plan your work so that you can avoid a big rush right before the final project deadline, and delegate different modules and responsibilities among your team members. Write this in terms of weekly deadlines.

I asked for permission to work alone due to time constraints. At this point I do not have exact schedule, but it should be something like this.

1.   Week Mar-15. Screen-scrape box office results. , start process book as blog, and create web site
2.   Week Mar-22. Get twitter counts data , Data analysis to check trends. Finalize visual design
3.   Week Mar-29. Coding main visualization
4.   Week Apr-5.   Get prototype ready
5. Week April 12 – Feedback re-factoring
6. Week April 19 – More coding improvements , + animation or secondary visualization for timeline
7. Week April 26 – finalize visualization, 2 minute video presentation

 Wow. That seems ambiguous but it should be fun to see what I can get done.