StockTwits Comments Analysis

Bingxin Chen bc1958@nyu.edu
Bihan Xiao bx339@nyu.edu
Zihan Liao zl1443@nyu.edu

Project page (on Github):
https://github.com/NYU-CS6313-SPRING2016/Group-8-StockTweets
Video: https://vimeo.com/167396667
Working demo: http://chronoresister.github.io/StockTweets-group-8/

What is the problem you want to solve and who has this problem?

www.stocktwits.com is a stock forum which has the similar pattern as Twitter, that people can add a "$" then follow an stock symbol(e.g: $AAPL) to comment a stock symbol, and also can using 2 tags "bullish" and "bearish" to give prediction. Giving a great amount of historical comment data on this website, we want to build a visualization tool, which can help psychologist and sociologist to analyze from how people talk about stock and how was the stock in reality. Some patterns in detail:

- What are the most popular stocks? Why are they popular?
- Based on these popular stocks, how do people predict stocks from price and trading volume?
- Based on these popular stocks, how do stocks reality feed back to people?

What are the driving analytical questions you want to be able to answer with your visualization?

- What are the top 50 popular stocks being talked about in a certain week?
- What are the key word in comments in a certain week?
- What are the stock price and trading volume changing of a certain stock in that week?
- How does people's comments about that stock change in that week?
- How does people's sentiment about that stock change in that week?
- What are the key word in comments for that stock, that week?
- What are the comments like? Provide some details.
- How do real data of stock(price, trading volume) and people's attitude(comments volume, sentiment) affect or reflect each other?

What does your data look like? Where does it come from? What real-world phenomena does it capture?

The data comes from StockTwits' API. The historical includes all comments on StockTwits in 2015, and we process the data in many JSONs, which we need in charts.
Here is the data table:

Attribute Name	Attribute Type	Description	Range	Derived
symbol	Categorical	The ticker symbol of the stock with an $ in the beginning	String	N
symbol_id	Categorical	Unique id for each symbol	String	N
symbol_exchange	Categorical	The platform on which the stock is trading on.	String	N
symbol_sector	Categorical	The industry sector the selected stock belongs to	String	N
symbol_title	Categorical	The name of the company	String	N
symbol_messageVolume_dailyCount	Quantative	The message volume of a selected stock in a day	[0, max]	Y
symbol_messageVolume_weeklyCount	Quantative	The message volume of a selected stock in a week	[0, max]	Y
weeklyKeywords	Categorical	The trending words of the selected week	String	Y
weeklyKeywords_count	Quantative	How many times the weeklyKeywords mentioned by users	[0, max]	Y
symbol_weeklyKeywords	Categorical	The trending words of the selected week related to the selected stock	String	Y
symbol_weeklyKeywords_count	Quantative	How many times the symbol_weeklyKeywords mentioned by users	[0, max]	Y
symbol_sentiment_bullish_count_daily	Quantative	How many messages about the selected stock is positive in a day.	[0, max]	Y

symbol_sentiment_bearish_daily	Quantative	How many messages about the selected stock is negative in a day.	[0, max]	Y
symbol_sentiment_value_daily	Quantative	The overall sentiment of the selected stock in a day	[-1, 1]	Y
symbol_sentiment_bullish_count_weekly	Quantative	How many messages about the selected stock is positive in a week.	[0, max]	Y
symbol_sentiment_bearish_weekly	Quantative	How many messages about the selected stock is negative in a week.	[0, max]	Y
symbol_sentiment_value_weekly	Quantative	The overall sentiment of the selected stock in a week	[-1, 1]	Y
time_price	Categorical	The date of the stock price.	String	N
time_message	Categorical	The time stamp of the message.	String	N
open_price	Quantative	The opening price of stock on a trading day.	[0, max]	N
close_price	Quantative	The closing price of stock on a trading day.	[0, max]	N
high_price	Quantative	The highest price of stock on a trading day.	[0, max]	N
low_price	Quantative	The lowest price of stock on a trading day.	[0, max]	N
time	Quantative	The subordinate of Visualization	[First date, Current]	N
volume_day	Quantative	The volume of message of a certain kind in a certain day.	[0, max]	Y
message_body	Categorical	The body of the message	String	N
message_sentiment	Categorical	The sentiment of that message	String	N

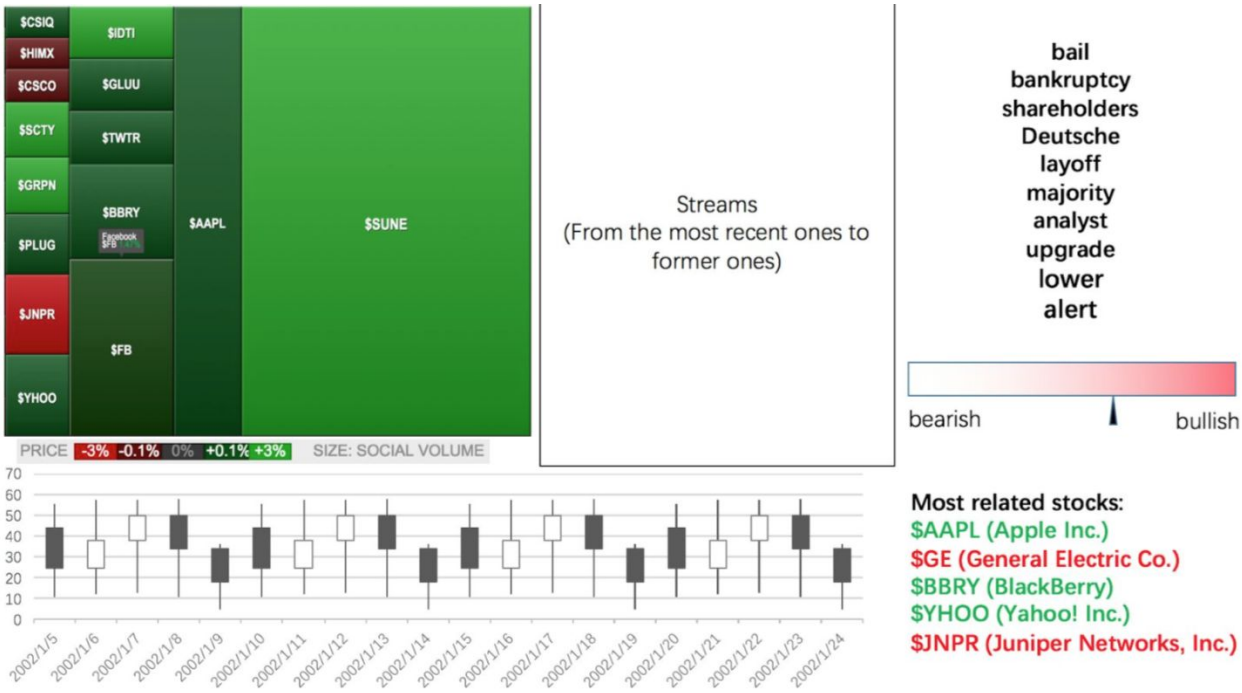
trending rank	Quantative	The rank of the stock in the trending list on a certain day.	[1, 30]	N
relatedStock	Categorical	Stocks symbol always mentioned together with the selected stock	String	Y
relatedStock_weekly Count	Quantative	How many times the relative stock and the selected stock are mentioned together	[0, max]	Y

What have others done to solve this or related problems?

Describe related works and explain how they are related to your work.
There is not too much visualization tools to analyze people's talking towards stocks. Related works of our tool commonly are similar as a forum or a monitor, or both.

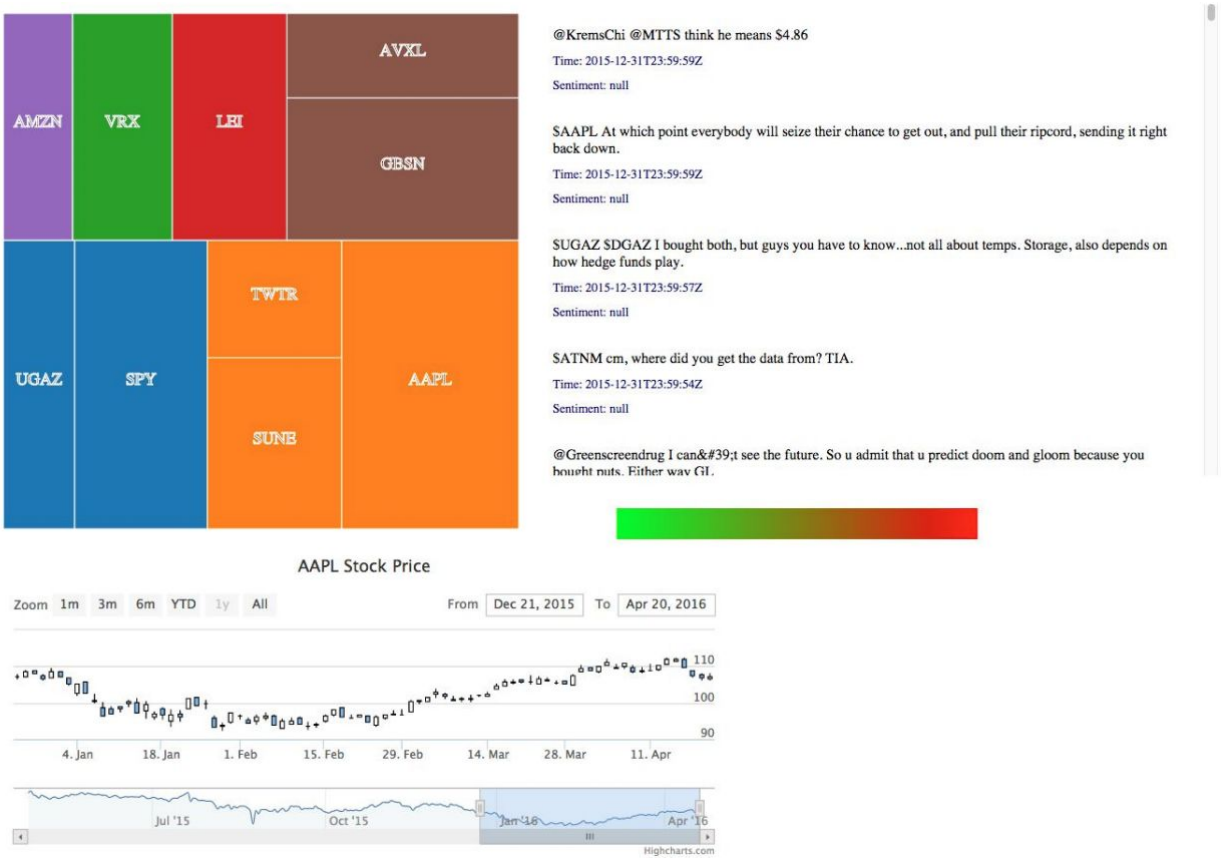
- www.stocktwits.com: StockTwits is a pure stock forum, and it is our data resource website. It focus on providing a Twitter-like mode with people to talk about stocks.
- http://www.thelion.com TheLion is a stock monitor with some news and a small forum. People mostly use this for monitor price and news.
- http://stockaholics.net Stockaholics also has monitor function, and it provides a thread mode to talk about stocks. Compared with StockTwits, this is better to give comments a theme, rather than see all the comments in parallel. The difference between this and StockTwits are just like BBS and Twitter.

Initial Mockup



- The top left corner is a treemap of top popular stocks in a period. The size is for their volume and color is for people's sentiment towards them, which shows as the legend below the treemap.
- The blank on the right of the treemap is using for rendering a stream of comments.
- The top right corner is a word cloud. It shows the key word of that period in form of d3.cloud
- The color bar below the word cloud is called sentiment bar. We process each comment and define "bullish" as 1 and "bearish" as 0, then add all the value together and divide by the number of comments. The result will be in range [0, 1], and the color bar is using hue to show the result.
- The chart in left bottom is stock price chart. You can click on the treemap to select a stock, then the chart shows price of that stock in that period.
- The list in right bottom is list of most relative stock. On stocktwits.com, each comment can add multi symbols, so we define that the symbols in one comment are relative. Click to select a stock on treemap and the list will render.

Project Update



This version just simply complete the mockup, so there is not much change from mockup.

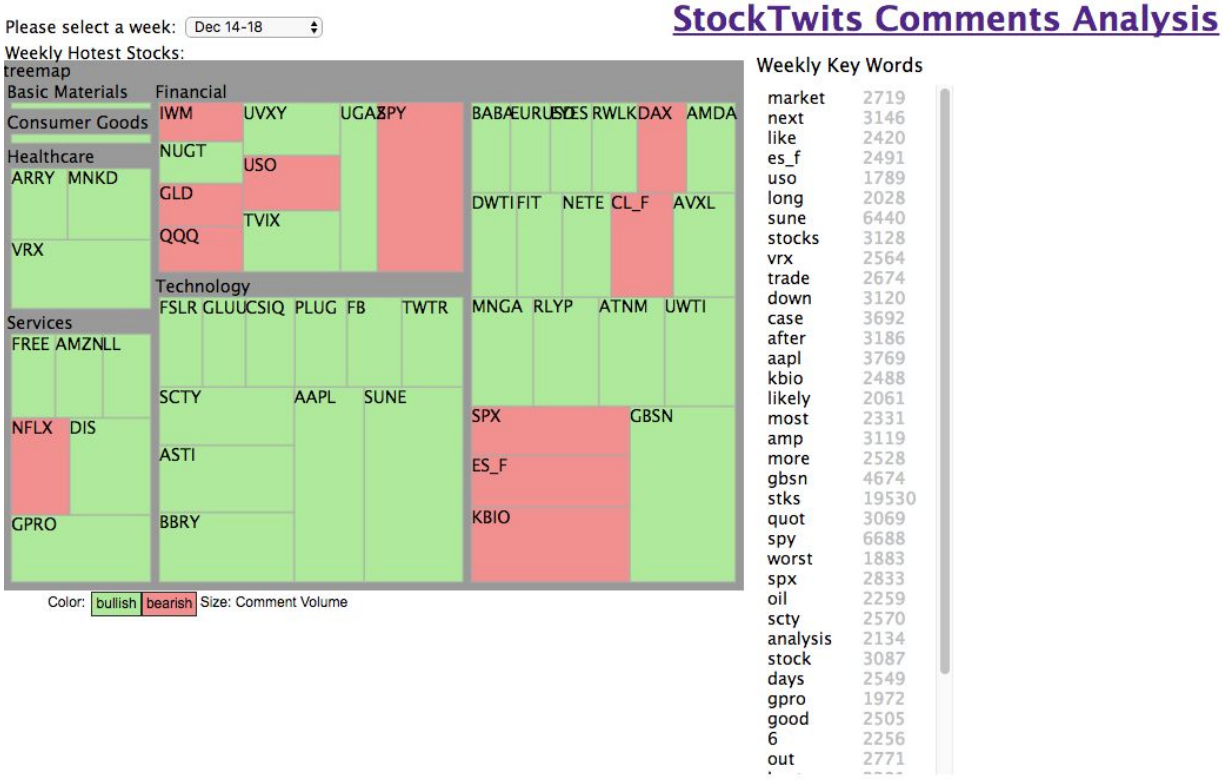
- The top left corner is the treemap. The size is for comment volume of each stock symbol, and colors are just random. Stocks are clickable to change the price chart below it.
- The left bottom is the chart of stock price and trading volume. It renders when a stock was selected by clicking on the treemap. The upper chart is stock price, which represented by many rectangles. Each rectangle represent stock price of one day. The lower chart is stock trading volume, and it is a simple line chart. There is also a slide bar to change the start time and end time.
- The top right corner is the stream list. Each line has three sub lines: message, time and sentiment.
- A color bar is on the image, but it has not been completed.

What did not work and improvement:

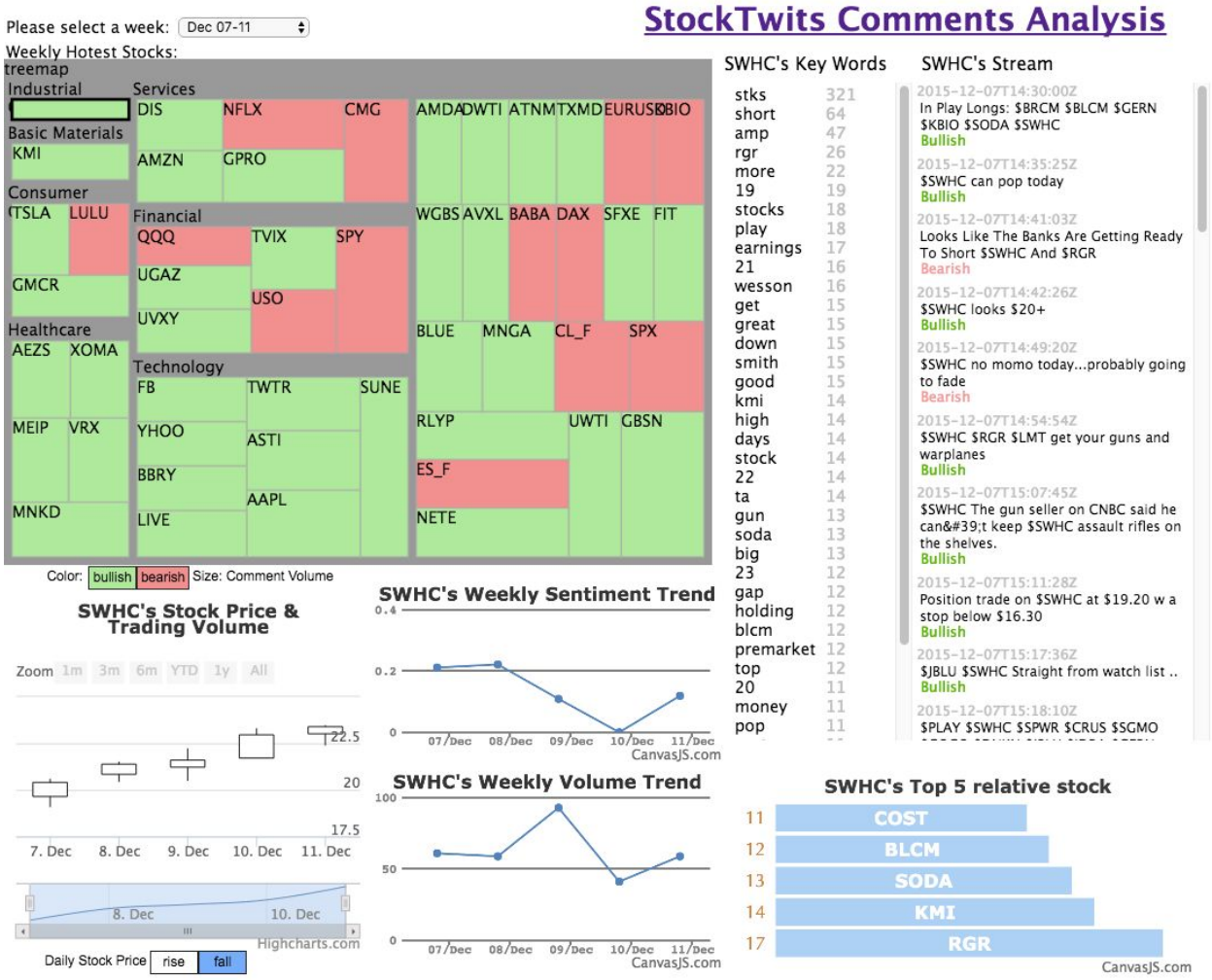
- The sentiment color bar is useless, because our time standard is a week, so we need to show how sentiment change in each day. Finally we change it into a sentiment line chart.

Final Visualization

Open the link of demo, you need to choose a week at first. The default week is from December 14th to 18th in 2015.



It only shows a treemap and a key word list, because other charts need a selected stock. Click the treemap twice, that first time to select a sector and second time to select a symbol, then other charts will be rendered.



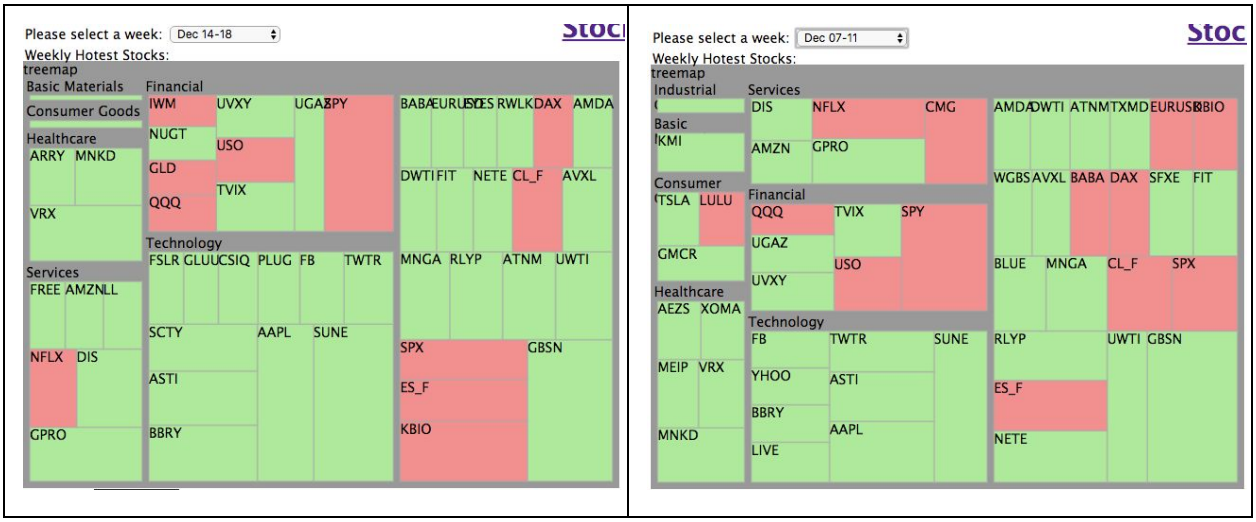
- There is a selection on the top left corner. It is to select a week of historical data.
- Treemap is on the top left corner. After selecting a week, the treemap shows all sectors of stocks, then click into a sector, then you can click to select a symbol.
- The key word list is on the right of treemap. If no stock is selected, it shows key word of that week, or it shows key word of the selected stock of that week. The list shows word and count.
- The stream list is on the right of word list. It shows when a stock is selected. The list shows some of comments of that stock in that week. It shows message, time and sentiment. Sentiment font color is same as the treemap.
- Price and trading volume chart is under the treemap. Similar as mentioned, it shows the selected stock's price and trading volume changing in that week.
- Sentiment trend and volume trend line chart is under the treemap, right of the price chart. They show the sentiment and volume changing in that week. The algorithm of sentiment value is the same as sentiment color bar in initial mockup, but the time is for each day.
- Relative stock bar chart is on the bottom right. It shows the top 5 relative stocks of the selected stock. The algorithm of relative stock is the same as initial mockup.

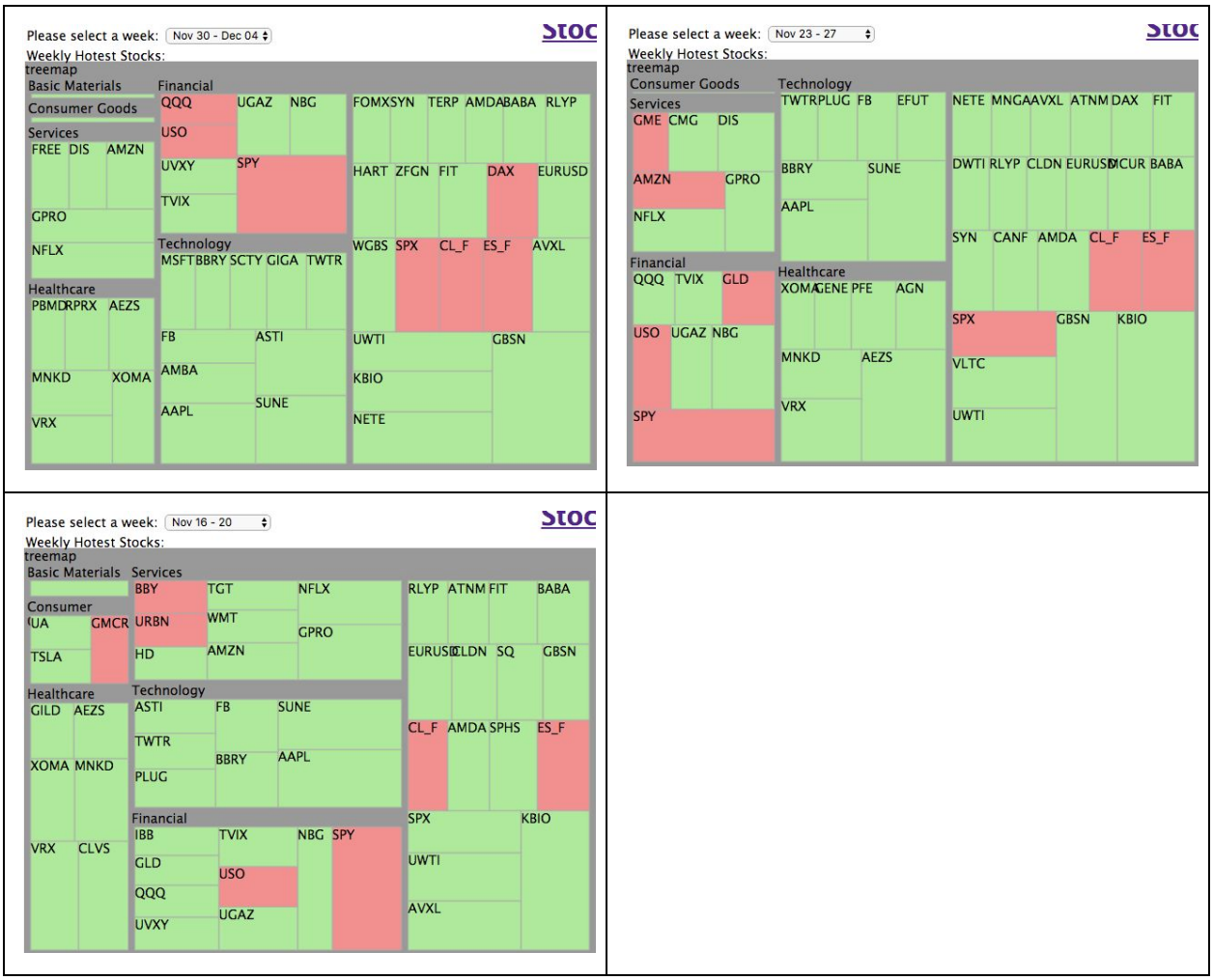
Improvement from mockup and update version:

- There is not a standard unit of time. To improve, we set two standard: week and day. Because the stock price API can only provide chart for at least 4 days, so we set a select entry as from Monday to Friday in each week(no market on weekends). Then after selecting a stock, we show data of this stock in each day of the week.
- Word cloud is not a good tool to express, but just a “cool” tool. Finally we give up it and change to a word list, which shows word and their counts.
- We change the sentiment color bar into sentiment line chart. This is necessary because our standard time unit is a week, then we entry each day of the week. So the line chart can reflect how sentiment change in the week better.
- Similarly, we add a comment volume line chart. Along with sentiment line chart, these two line chart can be used for comparing with stock price line chart and stock trading volume line chart. The change is really helpful for analyzing.

Data Analysis

- Compare each week’s treemap



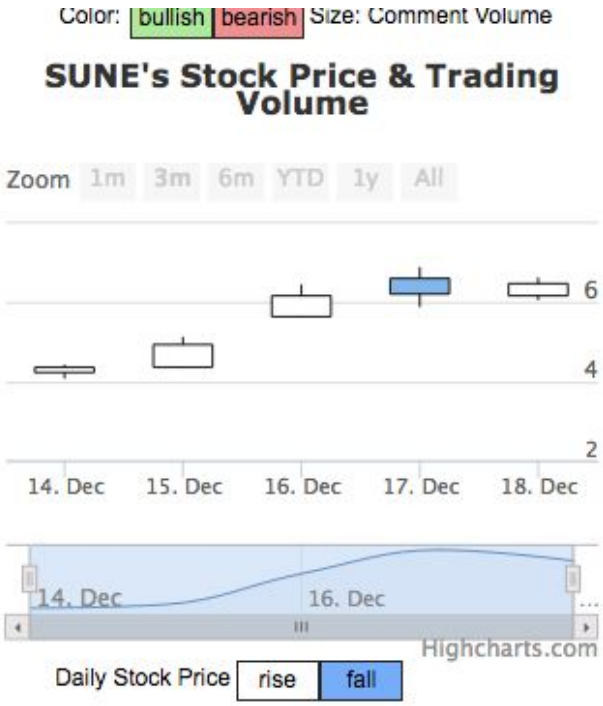


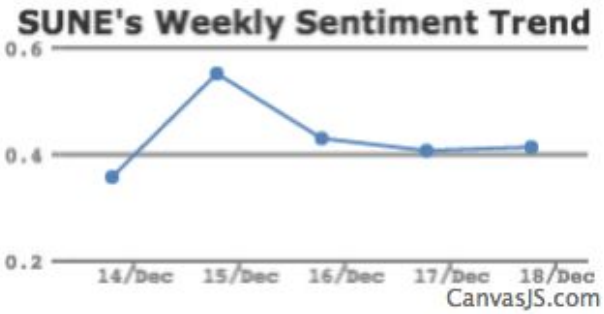
Obviously, green blocks are much more than red blocks. That means people always talk about what they think is bullish than bearish. Talking bullish can be comment for happiness, if the user bought a stock and it was going well. Also, talking bullish can be prediction, but talking bearish is less common in prediction because investors only focus on bullish.

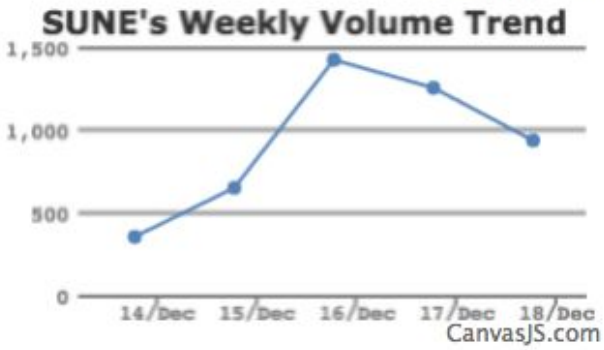
Besides, we found some stocks are always on the treemap, like AAPL, FB, GSBN… According to last paragraph, they should be usually in green, and that is right. Then we also found that, those always on the treemap or always be green are most in Technology sector. In our opinion, technology companies improve rapidly these years, so people have confidence on them.

- Compare the four line charts

For each selected stock, we provide 4 line charts: stock price, trading volume, sentiment trend and comment volume trend. We mainly compare stock price with sentiment trend, and trading volume with comment volume.



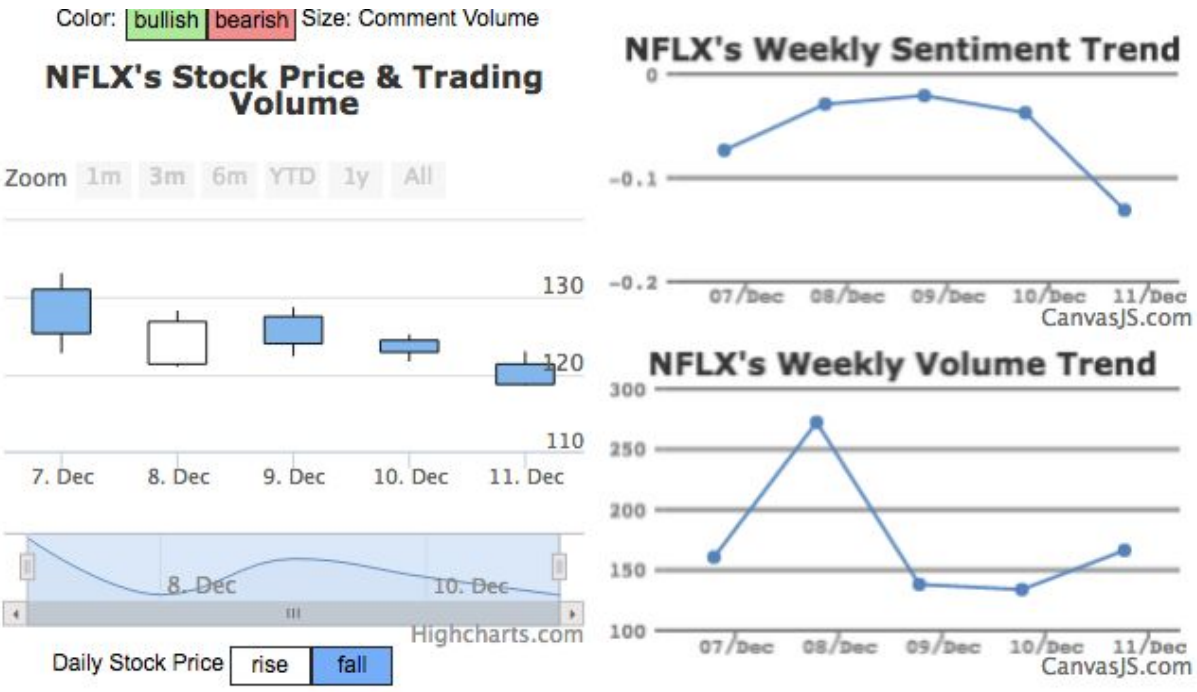




This is $SUNE in week of December 14th-18th. We believe this is the most common pattern.

Firstly compare stock price with sentiment trend: for the whole week, SUNE's price is improving, and the sentiment values are all above 0. Then see it in each day, on Dec 15 the price grows the most from open time to close time, and the sentiment of Dec 15 is the highest in this week. Dec 17 is the only day that open price is lower than close price, but the sentiment value of that day did not change much. We can say that people keep good prediction on it, and even if the price goes down on that day, since it is still higher than last day, so the day after's sentiment is still similar.

Then compare trading volume with comment volume, we found that the two line change similarly. It is easy to analyze that users on StockTwits are mostly investors, and those who traded for one stock have more possibility to make a comment about it.



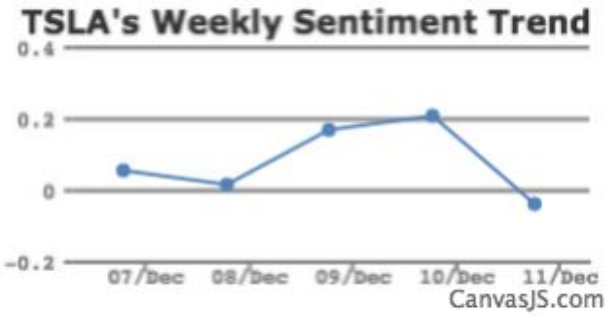
This is $NFLX from Dec 07 to 11. This is a counter pattern, which is not as common as the last one but also understandable.

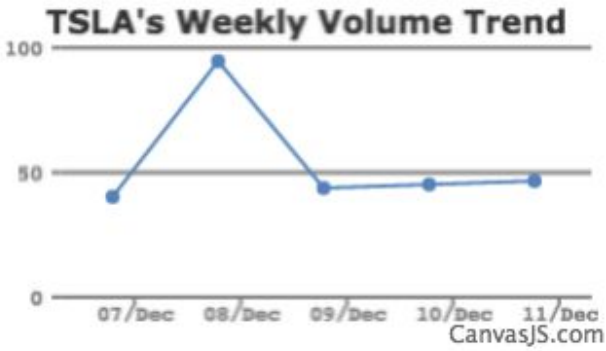
For the price, there is only one day Dec 8 that its close price was higher than open price. So it is easy to understand why the sentiment of whole week is below 0. However, the sentiment grows in Dec 08 and 09, maybe the price in Dec 08 gave people confidence.

The reason why we call it counter pattern is not for the price line and sentiment line, but the trading and comment volume. For the first glance on these two line, they are in counter: the trading volume goes down until Dec 8, then goes down again on Dec 9, but the comment volume grows on Dec 7 and goes down on Dec 8, then change a little.

In our opinion, this result reflects that trading has more information than comment, because trading has strategy. When encountering bearish, some investors may sell it quickly for less loss, but others may tend to keep it until the price goes back. However, comments on a forum just reflect people's concern about this stock, so there is always a great amount of comments when the stock price changed rapidly.

For this pattern, we found that a totally bullish week are easier to analyze.





This is TSLA in Dec 07 - 11. We take this as example to show some phenomena which are hard to explain.

First, for the sentiment trend line chart, it grows and become above 0 on Dec 09, but the stock price goes down a lot on that day.

Second, for the comment volume trend, it grows a lot on Dec 08, then return back on Dec 09, but there is no issue based on these 4 charts for what happened on Dec 08.

We must say, the only 5 day line charts and only forum comment is limited. There are always some visualized results hard to explain, but the two patterns above can be proved reasonable, with trying more data.

- Relative Stocks

There are also some patterns for relative stocks. We can both found it in bar chart and the key word list.

NFLX's Key Words

amzn	892
stks	83
short	71
fb	69
130	54
stock	47
out	44
market	40
aapl	40
like	37
money	37
goog	36
googl	35
more	34
quot	33
lol	28
long	28
netflix	27
good	25
stocks	25
spy	24
get	24
high	23
down	22
129	22
let	22
over	21
when	21
126	21

NFLX's Stream

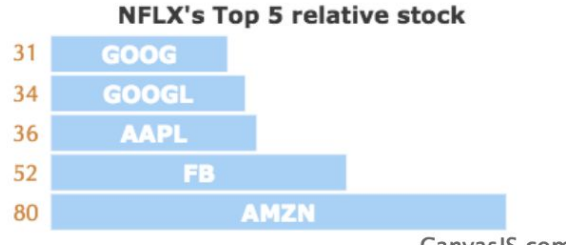
and Pirates $NFLX #Netflix http://stks.co/g3X0h
Bullish

2015-12-01T14:28:46Z
$NFLX Lord please forgive me for my sins. I promise to not gamble on a momo stock like this again. Please let us see $120 today. Amen.
Bearish

2015-12-01T14:32:40Z
$NFLX seems like a decent amount of us on here are short. Come on guys, let's hold hands and form a prayer circle
Bearish

2015-12-01T14:32:44Z
$NFLX it's a race out of yesterdays Bear trap !
Bullish

2015-12-01T14:34:01Z
$NFLX seems like a decent amount of us on here are bearish. Come on guys, let's hold hands and form a prayer circle
Bearish



AMZN's Key Words

stks	495
nflx	165
amazon	84
fb	60
spy	59
googl	44
aapl	38
quot	36
goog	36
stock	36
sales	33
msft	29
more	28
stocks	24
wmt	22
price	19
back	19
market	19
tgt	19
tsla	19
qqq	19
out	17
fang	16
good	16
50	15
700	15
like	14
down	14
baba	14
watch	14
friday	13
online	13
morning	13
calls	13

AMZN's Stream

$XLK $CMG I would not buy anything. Russia is about to have a WAR WITH ISLAMIC COUNTRY TURKEY. IM SELLIN
Bearish

2015-12-01T13:03:56Z
$AMBA $AMZN $MSFT $CMG $SPY wa with Islamic country turkey is immine according to Rueters and Russian new channels. Going to be rough
Bearish

2015-12-01T13:04:13Z
Early Movers $AMZN $ALL $PGR $YUN $JNJ $T $THO $CC $MFRM $BHP $VAL $DOW $AAL $BUD $DE $WYNN http://stks.co/h3XNr
Bullish

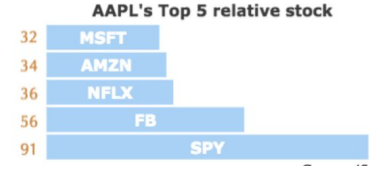
2015-12-01T13:11:36Z
$AMZN 50$ tablet umh; is it just an expensive mirror? Can you really do something with it?
Bearish

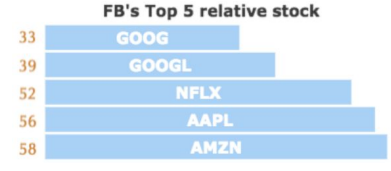
2015-12-01T13:27:06Z
$AMZN I would like it rip harder to upside premarket. I'm skeptical pricing action so far.
Bullish

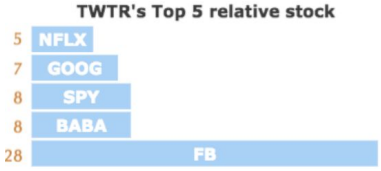
2015-12-01T13:39:44Z
$AMZN is a great example of how lon shallow bases work well when they #breakout. Great buy for #trendfollowing!
Bullish

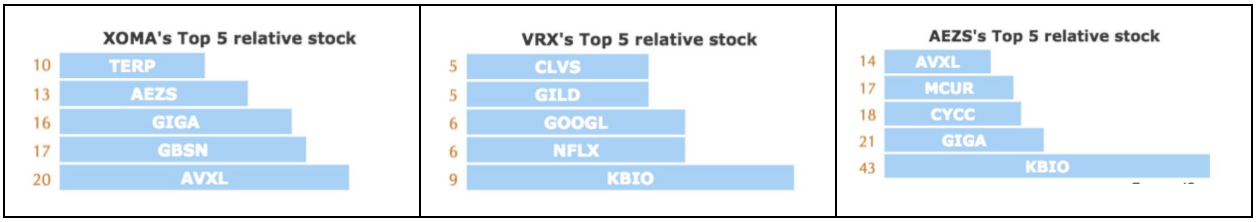
2015-12-01T13:45:06Z
$AMZN This thanksgiving I think the retailers have also stepped up on thei online orders winning back some of tl share back

Take NFLX as example, it is so obvious that NFLX and AMZN are greatly related. We believe that besides the stock issue, it is because these two companies are similar.









Also, stocks in same sector tend to be relative, especially in technology sector. For stocks in technology, 3 or 4 of top 5 relative stocks are in technologies, like row 1 in the table. For sector healthcare, it is common that 1 or 2 of top 5 relative stocks are in healthcare. For this phenomenon, besides stock reason, we think that technologie companies have more similarity and cooperation, but there are more competition for other sectors.

Limitations and Future Works

Limitations:

- Time selection is restricted. Only 5-day-week and single day can be choosed.
- There is no real-time data, so we cannot show the straight changing of comments
- The key word still need to be refined. We block some common meaningless word, but this result still remains meaningless word.
- Sentiment, key word would be better if using NLP(Natural Language Processing), but it is hard to implement since the forum comments seldom use formal English.

Next Steps:

- Change the time unit. The new version should be at least four days(or the price and trading volume can not be rendered) and at most about a month, by using a rolling bar to select.
- Add real-time stream, then compare with historical patterns. It will help a lot to analyze historical data.
- Try some NLP, hopefully there is powerful tools for cyber language and stock terminology.