Real-time Visualization of StockTwits

FINAL PROJECT DESCRIPTION

Project page (on Github): https://github.com/NYU-CS6313-SPRING2016/Group-7-StockTweets
Video: https://vimeo.com/167544405
Working demo: http://stocktwitsinfovis.herokuapp.com/infovis/

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

Stock market changes all the time, and investors are looking for every relevant information they can obtain to help them make best investing strategies. Public mood is a key feature of stock market which could be a potential indicator of market movement.

StockTwits (ST) is an online social media platform for sharing messages between investors, traders, and entrepreneurs. With more than 300,000 users, it provides a powerful platform for investors and analysts to analyze social mood of stock market. The large volume of messages created by ST users provide a plethora of market information.

However, such huge amount of raw data obscures people to quickly gain useful information of social mood. In this project, we want to visualize the stream of ST user messages, which could help stock investors and analysts to have a better understanding of social mood on stock market.

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

1. Market Overview: the whole market
    - What is the current whole market sentiment and trending of ST user messages? what is the volume?
    - Which sectors are user talking about most? Which stocks are users talking about most in these sector? What is the volume of the sentiment of these stocks?
    - Which stocks are users talking about most?
    - What are the words that ST users mentioned most on the market?
    - What are the contents of latest user messages? Are they bullish or bearish?
2. Market Details: selected stock or sector
    1) Sector

- What is the sentiment, volume and trending of messages in the selected sector?
- What does the sentiment look like in this sector?
- Which stocks are users talking about most in this sector? How much is the volume of these stocks? What does the sentiment look like?
- What are the words that ST users mentioned most on a sector?
- What are latest user messages in this sector?

2) Stock

- What is the sentiment, volume and trending of messages of the selected stock?
- How does current message volume change compared with past hours?
- What are the words that ST users mentioned most on the stock?
- What are latest user messages about it?

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

Our data comes from StockTwits API: http://stocktwits.com/developers/docs.
We constructed three tables based on the data fetched from the API.

Attribute name	**Attribute type**	**Attribute description**	**Value range/ Categories**	**If derived**
Stock				
stock_symbol	categorical	stock symbol	--	No
stock_title	categorical	company name	--	No
stock_sector	categorical	stock sectors	--	No
stock_id	categorical	stock id	--	No
stock_count	quantitative	number of messages mentioned the stock	[0, max]	Yes
Messages of selected stock/sector/overview				
messages_time	categorical	created time of messages	--	No
messages_volume	quantitative	number of messages	[0, max]	Yes
messages_sentiment	quantitative	sentiment index of messages	[0-100%]	Yes
Message				

message_id	categorical	message id	--	No
username	categorical	username of the message	--	No
user_id	categorical	user id	--	No
message_body	categorical	message body	--	No
message_sentiment	categorical	sentiment of the message	Neutral, Bullish, Bearish	Yes
message_time	categorical	Created time of the message	--	No
Word				
word_body	categorical	message body	--	No
word_count	quantitative	number of messages mentioned the word	[0, max]	Yes

What have others done to solve this or related problems?

There are several projects working on related problems and they have different data source, like http://tradethesentiment.com/ by using messages from Twitter. And some other projects like www.tradingview.com have used API from stocktwits, but they mainly focus on the stock price charts and haven't effectively visualize the setiment from uses in ST.

Initial Mockup

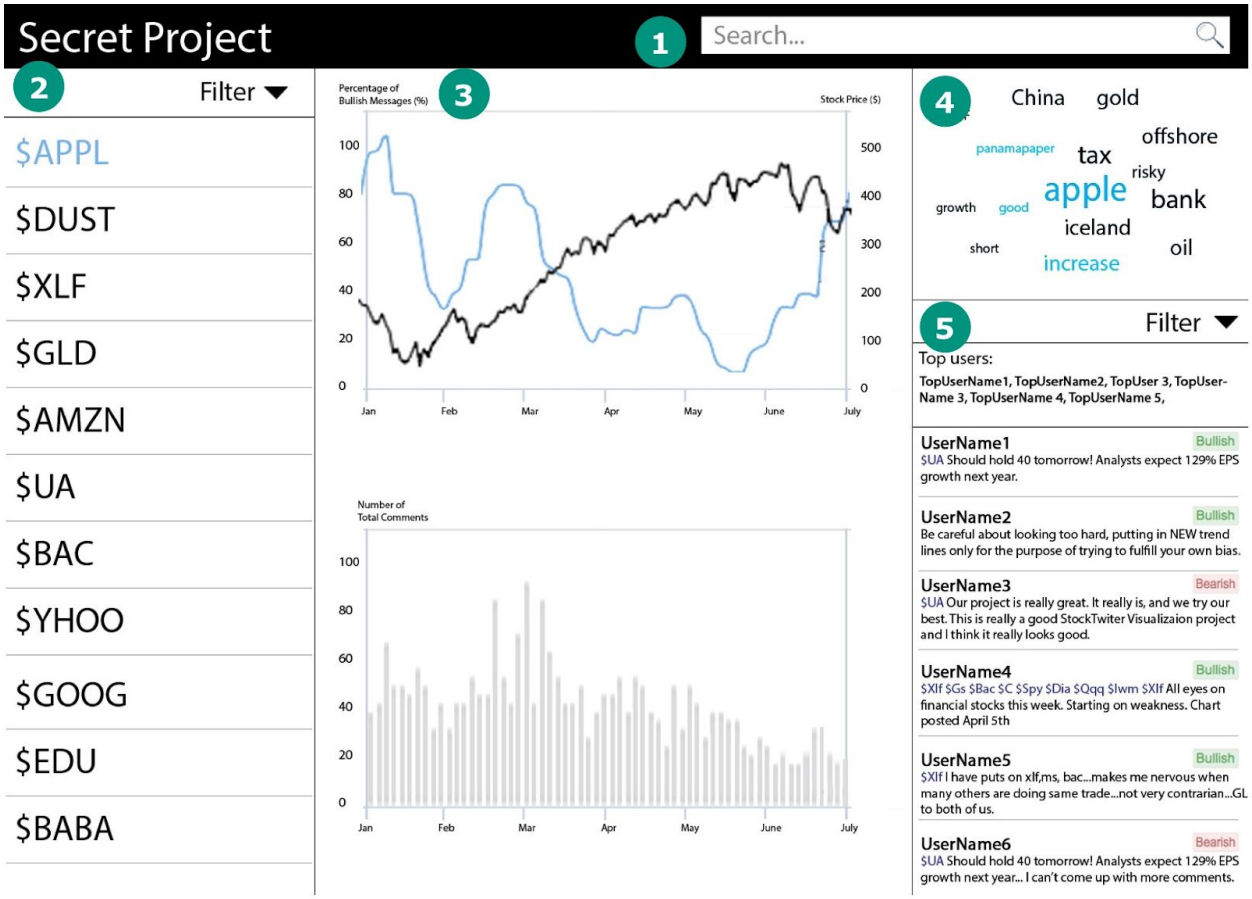
1. Web page and description

Page 1 - Overview

The overview page consists of 5 sections:

- Section 1 is a search bar which enables users to search for a specific stock.
- Section 2 is a list of stocks sorted by the number of messages on the stock. On top of this section, there is a filter which enables user to filter stocks by sector(industry). When hovering a stock on the list, the viz will provide user details of the corresponding stock by a tooltip, and the stock is highlighted between both the list and the treemap. When user clicks on a stock, the page will switch to page 2, which provides more detailed information of selected stock.
- Section 3 is a treemap. The color hue of red and green represents negative and positive change of stock price, luminance represents the percentage of the change, and size represents the number of messages shared by ST users about the stock. When hovering on the list, the viz will provide details of the corresponding stock by tooltip. When user clicks on a stock, the page will switch to page 2.
- Section 4 is a keyword cloud displaying the most mentioned keywords in all messages. The size of the keyword encodes the frequency of each word mentioned by ST users.

- Section 5 is a real-time display of ST users messages. User can filter the messages from specific ST users, like suggested ST users. There is a tag on the right side of each message, which provides the bullish/bearish information.
- Page 2 - Details of The Selected Stock



The stock page also consists of 5 sections.

- Section 1 is the search bar with same function as overview page.
- Section 2 is the stock list with same function as overview page, and the current selected stock is highlighted.
- Section 3 consists of 2 parts. The upper parts is a line chart, with blue line encoding percentage change of bullish/bearish messages of the stock (y-axis on the left), and the black line representing stock price change over time (y-axis on the right). The bottom part is a bar chart, which encodes the total number of messages on the stock in different time period.
- Section 4 is a keyword cloud. Compared with the overview page, the keyword cloud here is from all messages of the selected stock, and the size of keyword encodes the frequency of each word mentioned by ST users of the stock, the luminance of keyword encodes the relevance of the keyword with the stock.
- Section 5 is a real-time display of ST users messages about selected stock, and it also provides the names of 5 top users who share most messages on selected messages.

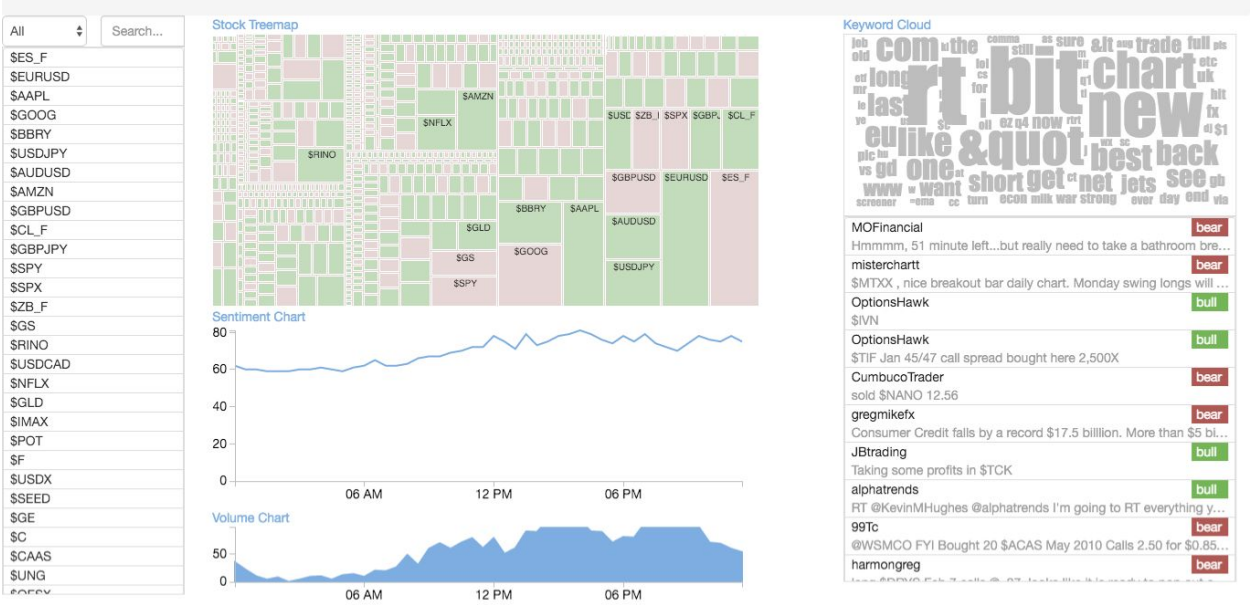
2. What did not work and how we changed to improve it.

- The space is not utilized effectively and information is duplicated. We displayed the treemap, line chart and area chart in a single page, rendering all the charts in a single page is easier for user to get different information at the same time rather than having to switch between pages.
- We changed our treemap color hue channel from encoding stock price to user sentiment. This is because 1) we can't get real time stock price (real time stock service is really expensive), 2) we want to focus more on real time display, 3) we think user sentiment is an interest point.
- Bar chart focus more on individual value and their comparison, however, we want to focus on the total volume. As a result, we changed our bar chart to area chart when encoding user message volume.

These changes are realized in the next section -- project update.

Project Update

1. Web page and description



There are 6 parts in the page: stocklist, treemap, line chart, area chart, keyword cloud and messagelist.

1) Stocklist: the list displays top 50 stocks by message volume. It also provides filter function by sector and a search function.
2) Treemap: color hue encodes sentiment category(bullish/bearish), size encode message volume. Use real-time data.
3) Line chart encodes real-time stock sentiment (proportion of bullish messages) within 24 hours.
4) Area chart encodes user message volume within 24 hours.

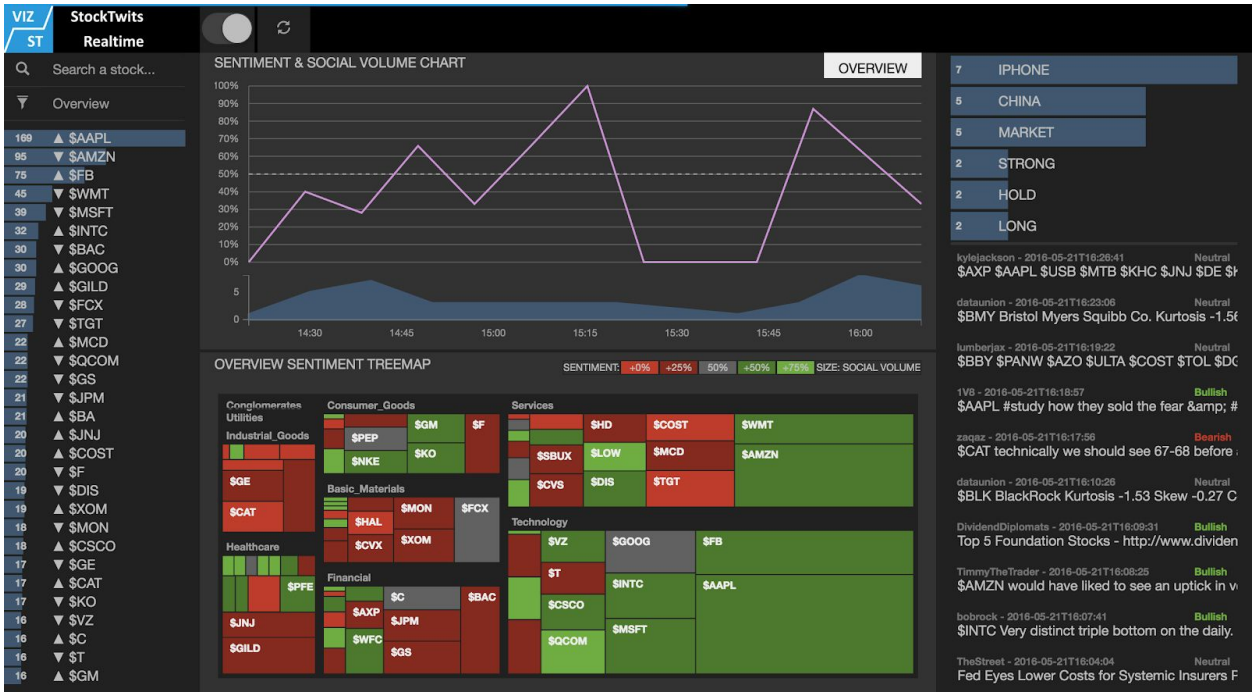
5) Keyword cloud: size encodes frequency of each word mentioned during the past 24 hours.
6) messagelist: real-time display of user messages and the corresponding sentiment.

By default, the treemap, line chart, area chart and keyword cloud will display the overall data of most mentioned stocks (top 50 for now) in StockTwits. When user clicks on a stock in stocklist or treemap, the treemap will highlight the corresponding stock, and line chart, area chart, keyword cloud and message list will update the information for the selected stock.

2. What did not work and how we changed to improve it.
   - We believe line chart and area chart are more important than tree map, because they provide more detailed information of the market, sector and stocks. As a result, we substitute the position of line and bar chart with tree map.
   - List on the left can only provide the sorted information of stock, which is not a good use of space. We added a bar chart and encoded the actual number of message volume to the list.
   - Word cloud is not a good visualization tool, it takes up too much space and only supports crude sorts of textual analysis. As a result, we changed it to keyword list.

These changes are realized in the next section -- final visualization.

Final Visualization



1. The final visualization is a real-time monitor of StockTwits forum, we display data within the latest 2 hours. Users can click on the button on the upper left to turn off the real-time display and refresh manually. By default the auto-refreshing is turned on and animations happens in all lists, charts and treemap. We learned from lecture that animation may be annoying and divert users' attention, so we optimize our algorithm and make such animations smoothly.

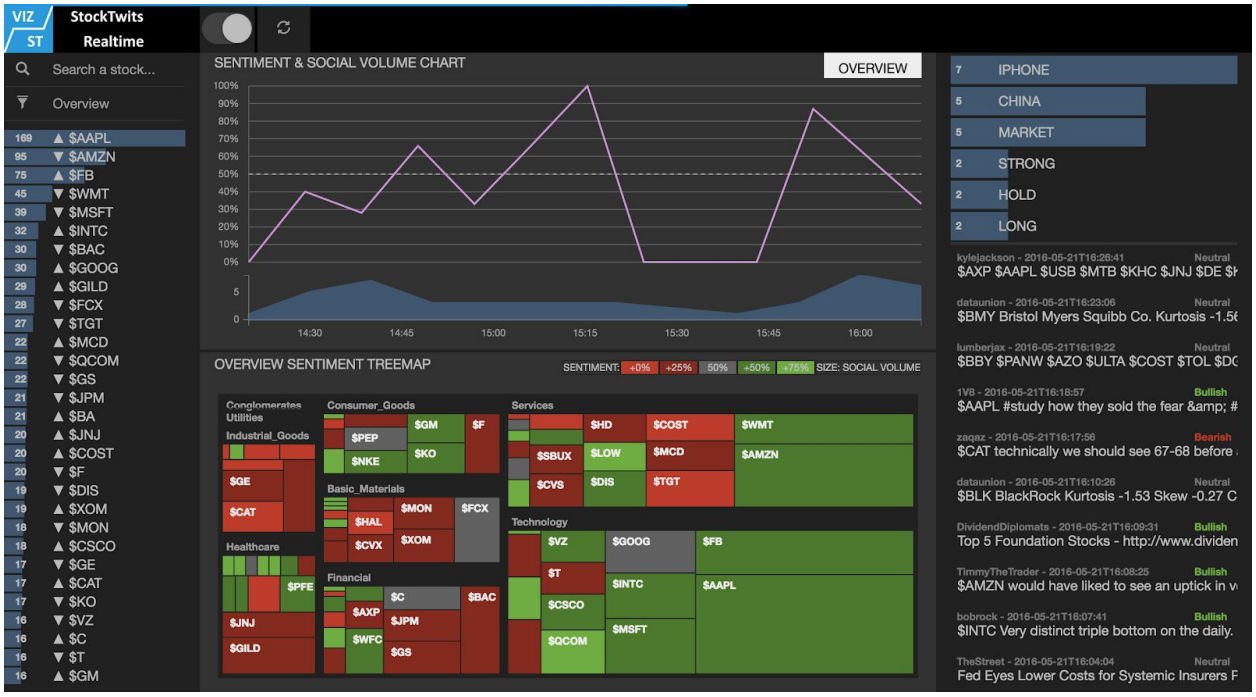
2. The final visualization takes up of 6 sections.
    1) Stocklist: displays top 30 stocks by message volume. Along with the list, we encode a bar chart as a background, with the horizontal position representing message volume and mark the exact number on the left. There are also filter and search function. Users can search stocks by symbols, and filter stocks by sectors.
    2) Treemap: color hue encodes sentiment category(bullish/bearish), color luminance encodes the intensity of sentiment, cell size encodes message volume. When user clicks on a stock, the tree map will zoom in to show the sector view.
    3) Line: chart encodes real-time stock sentiment (proportion of bullish messages) within the past 2 hours.
    4) Area chart: encodes user message volume within the past 2 hours.
    5) Keyword list: encodes keyword frequency of user messages in the bar.
    6) Message list: displays real-time user messages and the corresponding sentiment. A bar is encoded in each line of the message list to display sentiment

analysis of this message. When clicking on an message, there will be a pop up to show details of the message.

3. There is linking between stocklist and treemap.
4. When hovering on stocklist, treemap, line chart and bar chart, there will be a tooltip to show details of the stock.
5. When a sector or stock is selected in the stock list, the whole page will refresh to show the corresponding data.

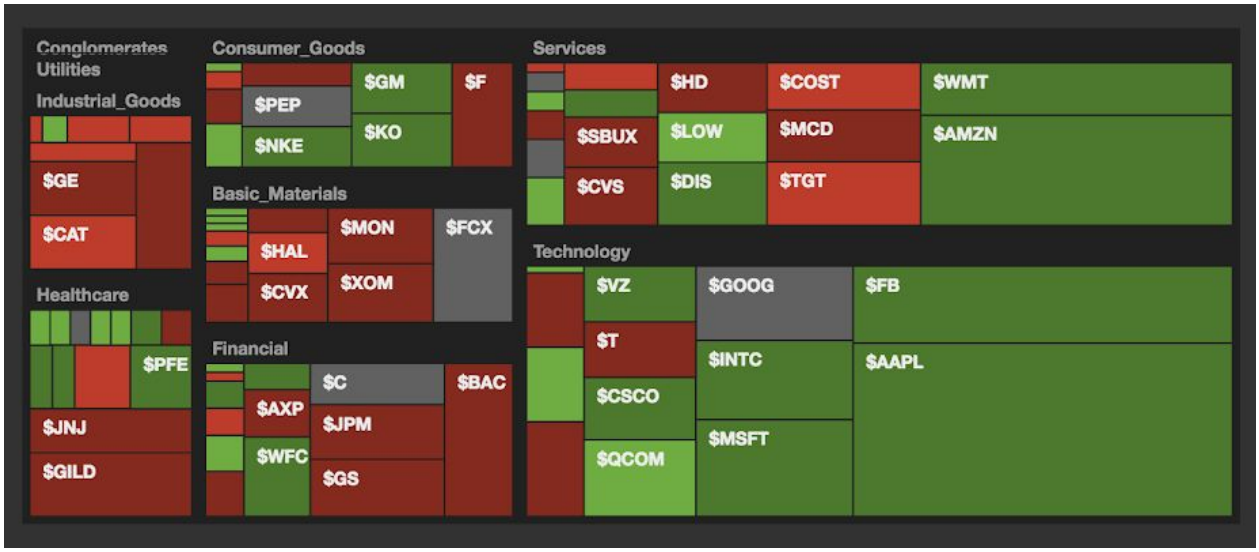
Data Analysis

Let's assume at noon on a weekday when stock exchange is taking a break, some investors are exchanging information on StockTwits on the performance of the market in the morning and making comments on the possible performance in the afternoon.
First look at the overview of the whole market messages.

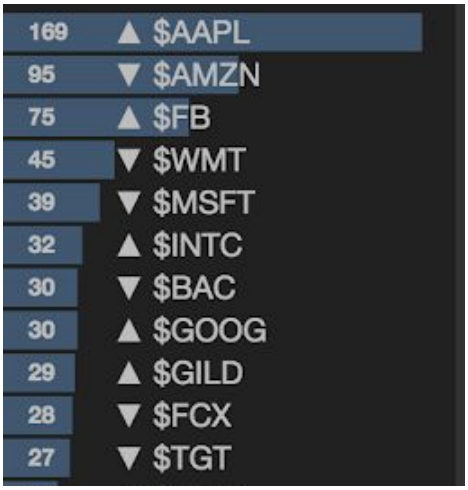


Investors can see from the stocklist that stock $AAPL has the most volume of users comments, the triangle indicates the volume increased during the past hours. From the line and area chart, investors can have an overview of trending of user sentiment and message volume during the past hours. They can also see the main focus of StockTwits users from the word list, and details of user messages in the message list.

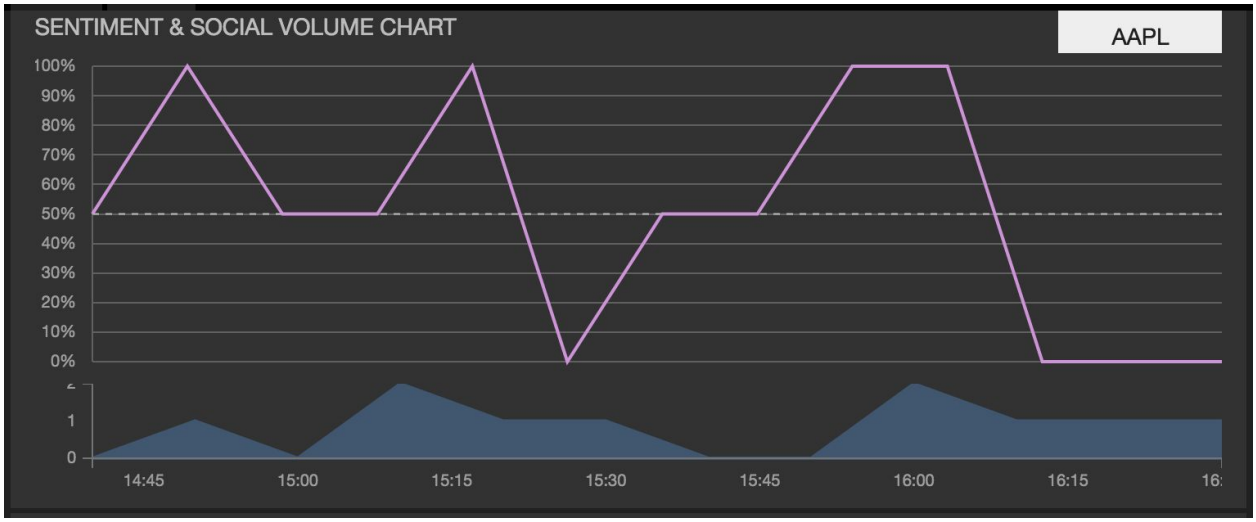
Then they may ask what about other information? Which sector has impacted the market most? Probably it is the sector that users are commenting most. Let's take a look at the treemap:



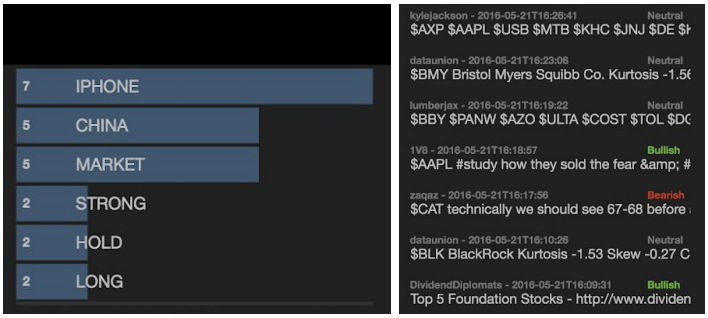
We find that technology sector has the most volume of user messages. This is an indication that people are more interested in technology sector, probably stock prices will have a change in this sector. But how will it change? Are people expecting rise or fall for these sectors? We can tell from the color of the treemap which encodes the sentiment of user messages. The treemap is mostly green in technology, so maybe stock index focusing more on this sector will go up in the following hours. What's more, we can see that, in tech sector, Apple, Facebook and Microsoft are attracting more user attention compared with other stocks in this sector. So investors may want to invest more in these stocks in order to take advantages of tech sector. But we can only tell the current message volume from the treemap, how did this volume change during the past hours? Let's take a look at the stock list on the left.



We can tell from the triangle in the list, which indicates the message volume change during the past hours. The message volume of Apple and Facebook are increasing, does this mean that trading volume will increase? We are not sure, but let's select the specific stock first, e.g. Apple, we can see the details of a single stock:



From the line chart and area chart, we can see the trending of message sentiment and volume during the past hours. We can tell whether it's vibrating or trending upwards or downwards, this gives investor an impression of users' mood on this stock, which could help he/she making decisions. Furthermore, investors may want to see the details of user messages. Let's take a look at the word list and message list on the right.



From the word list, we can tell that words "IPHONE", "CHINA" and "MARKET" are most mentioned. This could give investors a clue of the StockTwits users' focuses. Investors can also see the details of real time messages and sentiments as well.

In general, investors can get real time information from StockTwits user messages, which is interesting and probably helpful for them to make decisions on investments.

Limitations and Future Works

1. Due to the limitation of the free version of heroku, we cannot buffer enough data to demonstrate our visualization in different time frame, such as 24 hours, a week and a month. Provided a much powerful platform, we can visualize more time frame views and provide summaries of historical StockTwits messages, which will give us more insight into StockTwits.
2. Currently we don't have access to real time stock price. If could, we would like to add the trend line of stock price to our line and area chart with which we could make a comparison between stock price and user sentiment, as well as message volume. Hopefully we could find some interesting correlations between these features.

ACKNOWLEDGEMENT

Thank you Professor Bertini, you are an amazing professor. We have learned a lot in this class, not only the wonderful visualization tools, but also from your personality and your way of doing things: to continuously ask questions to ourselves, to comment with good reasoning, not to satisfy with things we have done, but to look for more alternatives.

Thank you Anshul for your help on StockTwits visualization, especially when you were busy with your paper publication. Your suggestions on our project are really helpful and valuable.

Thank you Cristian, the d3 lab tutorial is fantastic. Thank you for the nice suggestions you gave on our project.

Thanks to class CS6313 Information Visualization, the three of us can work closely with each other, we treasure the days and nights we spent together, the amazing corporation we had and will have.