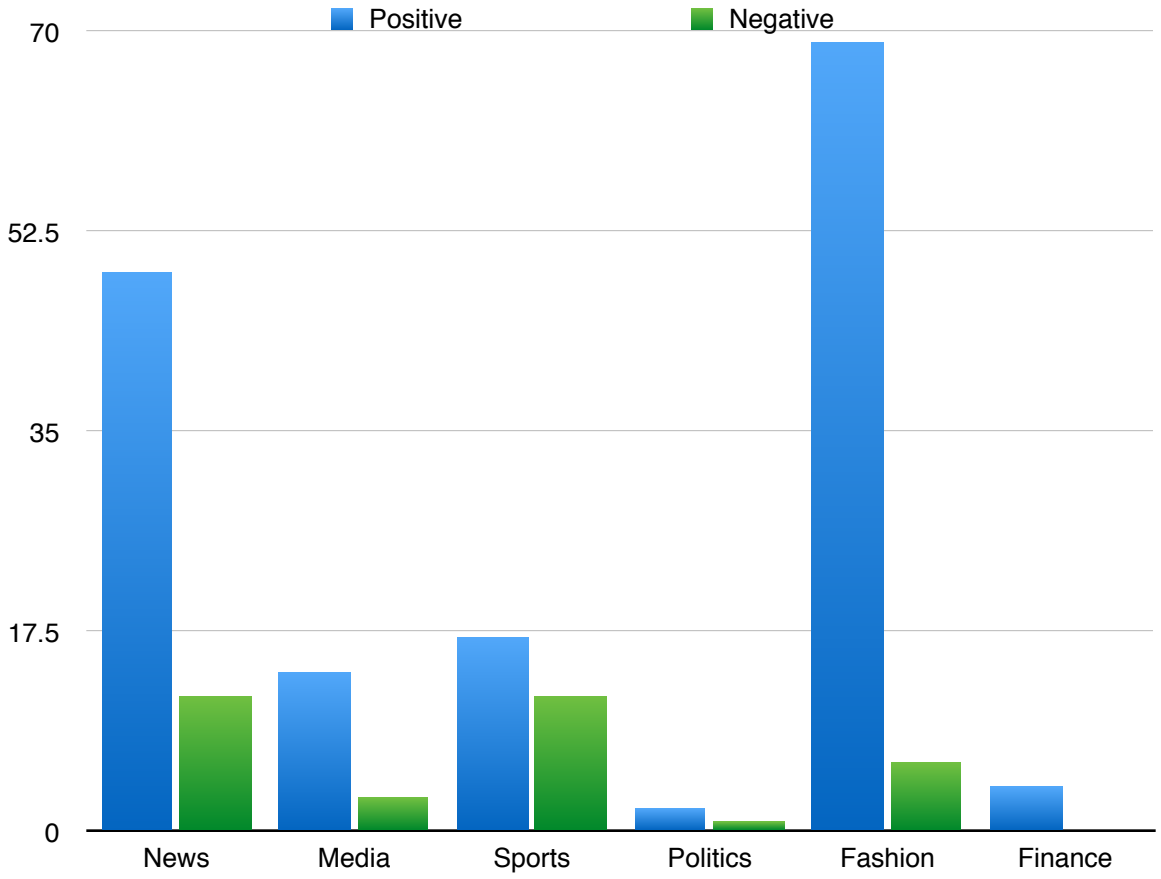
# Sentiment Visualization and Analysis

Wendan Kang (wk2269), Enrui Liao (el2756)

In Assignment 3, we choose two different keywords, using ourself data, to do sentiment analysis using EMR by Amazon. Sentiment Analysis is the process of detecting the contextual polarity of text. In other words, it determines whether a piece of writing is positive, negative or neutral. Out topic list is news, media, sports, politics, fashion, finance.

**FIRSTLY, WE CHOSE WORD "LOVE".**
We can easily see, in this keyword "love", no matter in which topic, most of them are analyzed as "positive". And it didn't change much as time goes. This information is telling us that "love" almost always means "positive". But it might get negative in some situations, such as "don't act like its a bad thing to fall in love with me."

We used dataset contains 20,000 tweets which contain "love". Result is listed below:



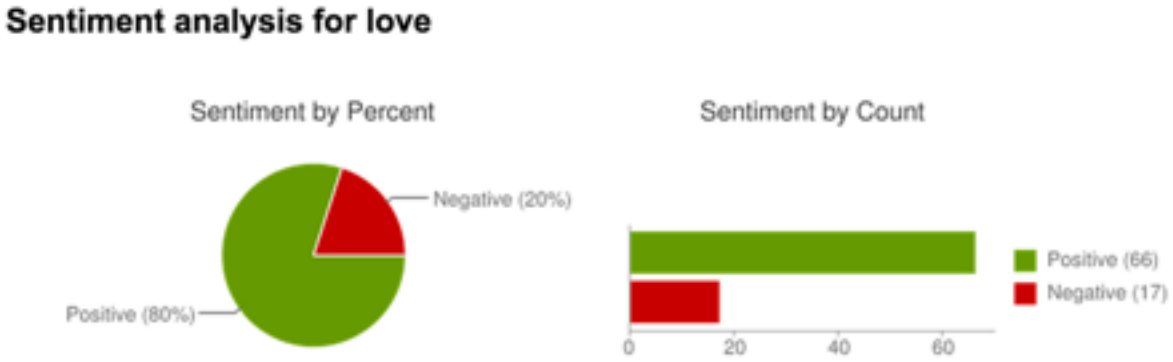

We can tell from the result that "love" get involved with "news" and "fashion" topics more frequently than other topics.

And in contrast, we use different sentimental tools online:

1. http://www.csc.ncsu.edu/faculty/healey/tweet_viz/tweet_app/


The right part of coordinate graph represents positive, left part represents negative. We can see that the result is same with ours. Most tweets about “love” are positive.

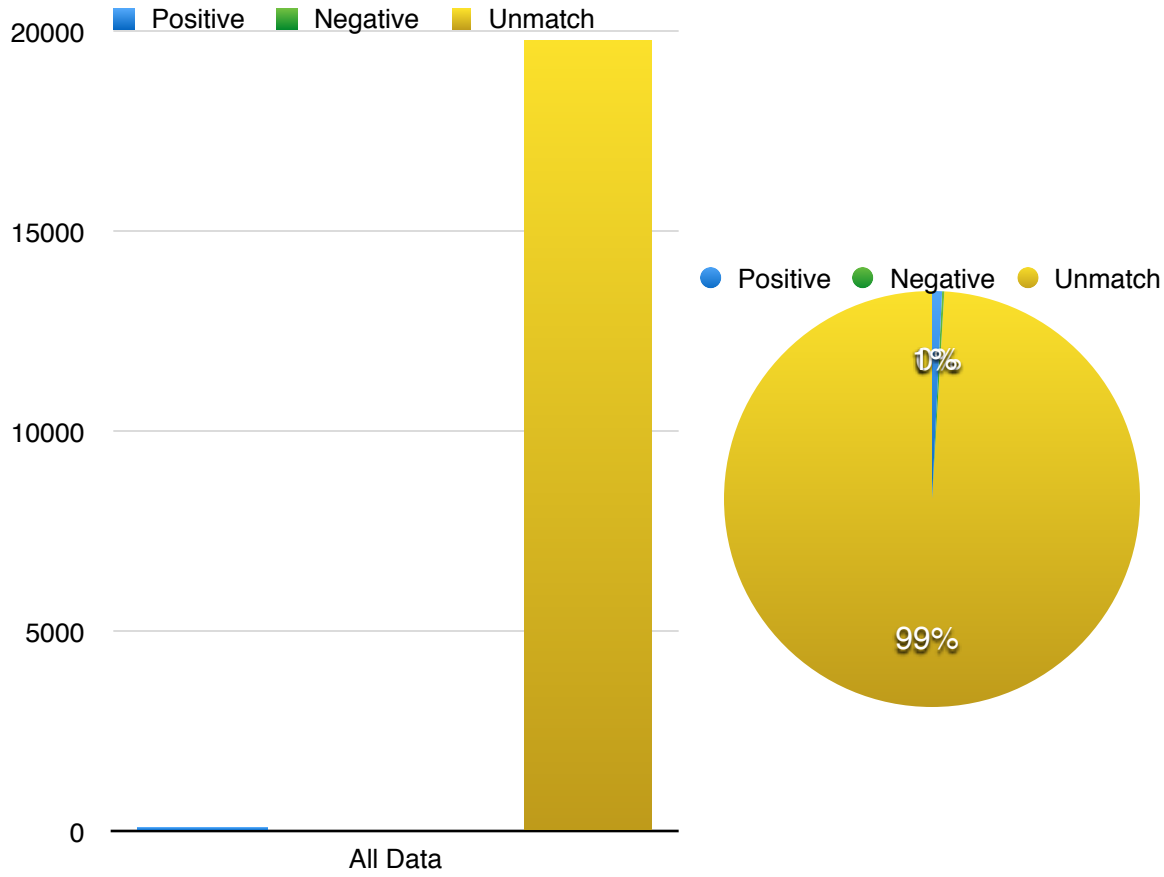
2. http://www.sentiment140.com/




By using this tool, the trend of results are same. But it shows negative results is more than ours. We think that depends on which kind of classifier we are using. Different classifier will cause different sentiment results.

Overall, our result is on the right track. And it is agreed by other results from different tools.

The charts over all data about love are listed below:




We can see that much of the tweets are classified as “Unmatch”. There might be several reasons for that, such as which classifier we are using, and the quality of tweets.