

MA415 Final Project

Nikita Pandya

12/12/2017

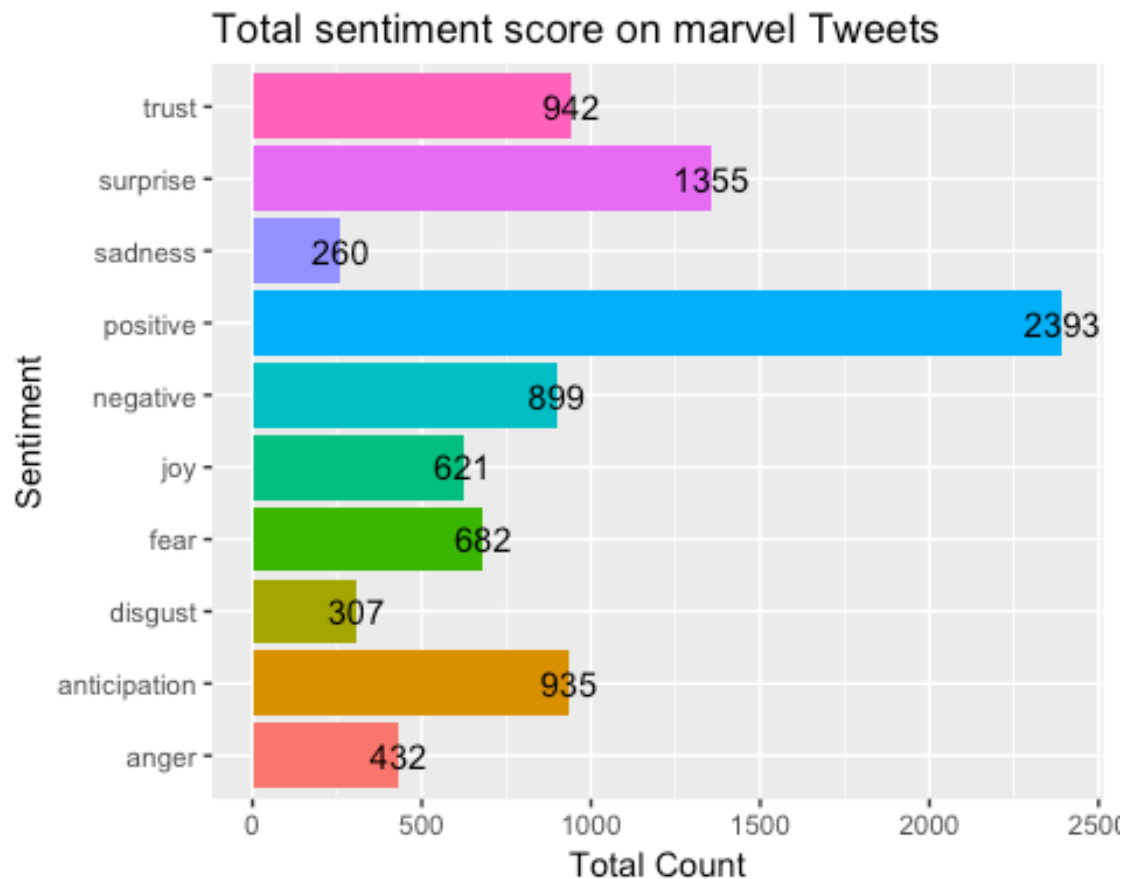
Twitter Sentiment Analysis: Marvel

I choose the brand Marvel to perform a sentiment analysis via twitter. I choose Marvel because last week they released a trailer for their next Avengers movie: "infinity war" and "Black Panther" and Netflix released revealed the season 2 premier date for Marvel's Jessica Jones. All the data was collected from twitter by using their API functions "searchTwitter" and userTimeline. I gathered tweets by using the keyword "#marvel" and by searching the official @Marvel twitter page. After gathering the tweets, I cleaned them by making all words lower case, removing punctuations and digits, removing retweets "RTs" and also removing usernames. This way the sentiment analysis and the word cloud would be as accurate as possible. The markdown contains the following analysis, plots and tables:

- Sentiment bar chart - breaks down tweets into different emotions
- Sentiment line plot - shows overall sentiment trends from the past week
- Top hashtags associated with the brand
- User statistics
 - Map to show where the people are tweeting about Marvel from
 - Histogram: tweets per hour
 - Table: top tweeters of hashtag
- Word cloud
- Emoji analysis

Note: All the data in this write-up was collected on December 12th. Plots and tables will vary slightly when recreated since new tweets will be gathered and cleaned for analysis.

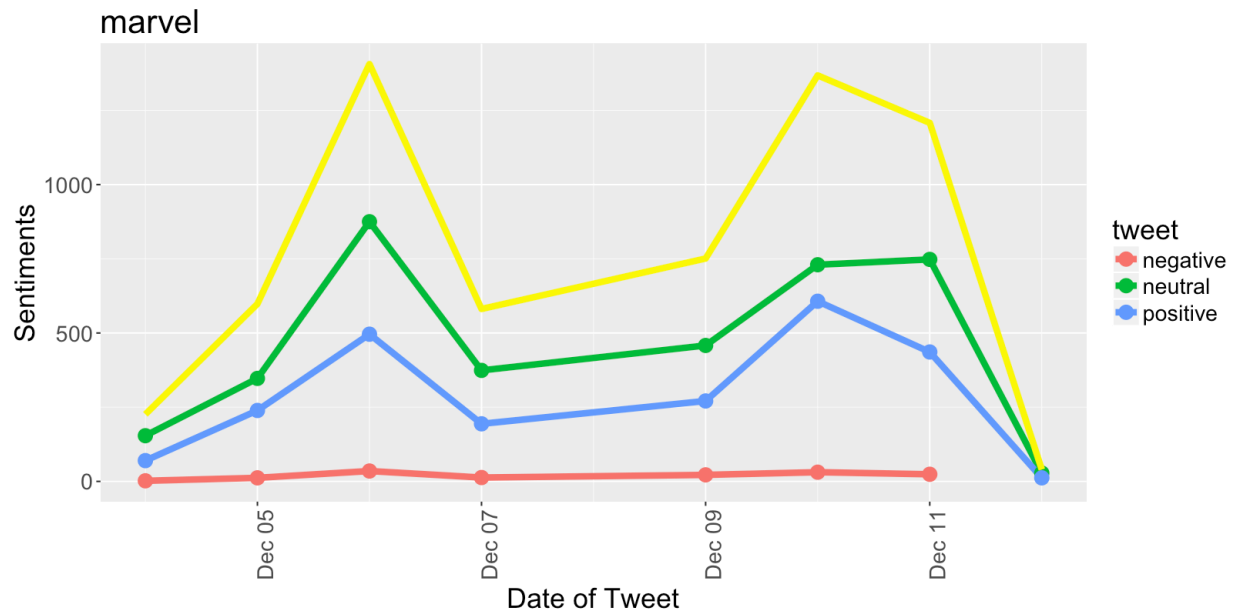
Sentiment Bar Chart



The overall sentiment towards Marvel seems to be positive, followed by surprise, trust and anticipation. The barchart was created using the "Syuzhet Package". This package includes a function called `get_nrc_sentiment()` which takes in the tweets (in the form of a character vector) and outputs a data frame where each row represents a sentence from the original file. The columns include one for each emotion type as well as a positive or negative valence. The ten columns are as follows: "anger", "anticipation", "disgust", "fear", "joy", "sadness", "surprise", "trust", "negative" and "positive." After attaining the sentiments, I use `ggplot` to create a bar plot to visualize the sentiment and the count.

Thus, we can conclude that overall people have had a positive reaction to Marvel's new movie trailers. People are also surprised and in anticipation of the movie. This positive reaction indicates that the movie upon release will be successful.

Line Plot



After gathering and cleaning the data, the function stores the collected tweets in a "marvel_stack.csv" file

```
##
text
## 1 RT @Todd_McFarlane: Some of the PERKS of being in the #comicbook busin
ess. A high grade VENOM variant cover. I helped co-create the charact...
## 2 RT @Le_sigh20: Decking the halls with @Marvel @MarvelStudios Christmas
Loki. Can't have a real tree since my dog would eat it, so this is t...
## 3
marvel invisible girl naked 100 porn videos https://t.co/EFYlqDgHyg
## 4 RT @ComicBook: Its nostalgia overload in this fanmade recreation of the
#AvengersInfinityWar trailer using only '90s #Marvel cartoons: \n\nht...
## 5
#Marvel's #AntManandTheWasp to be a rom com!?! \x
ed\xa0\xbd\xed\x81\xed\xa0\xbe\xed\x4\xa3 https://t.co/o6EtmVlzRj
## 6 @CaptainFrugal @Marvel if cebulski even remotely considers this statem
ent acceptable?\ni say let the dying beast of... https://t.co/BNDnzXnfsn
```

The function `score.sentiment` then evaluates the tweets and determines if their sentiment is positive, negative and neutral. This is done by scanning two files one filled with [positive words](#) and the other with [negative words](#). These scores are then saved in a "marvel_opin.csv" file. Then using `ggplot`, the scores are visualized on a line plot.

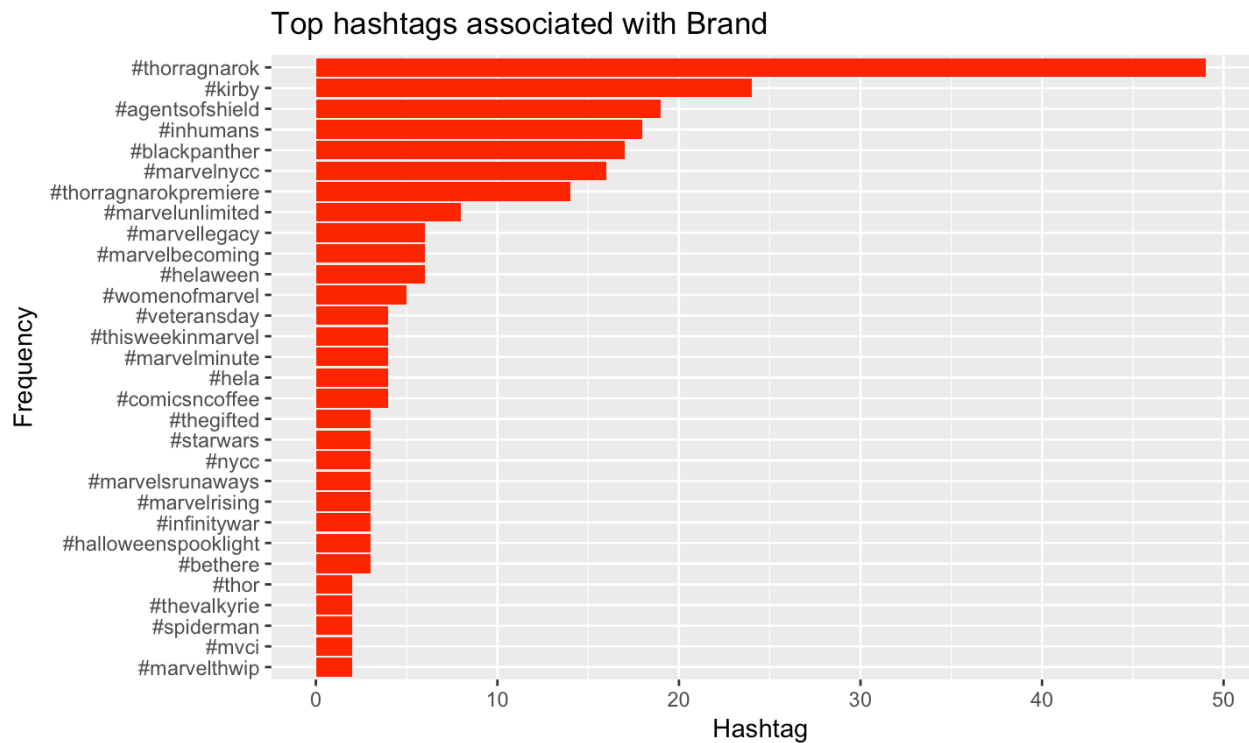
The majority of the tweets last week seem to have a neutral sentiment followed by a positive. There was a spike in the numbers of tweets with the keyword "Marvel" on the sixth and again on the tenth, around when the trailers were released. Thus, we can conclude that the

majority of people have had a positive reaction to Marvel's new Avenger's Infinity War movie trailer and the announcement of Jessica Jones season 2.

Tweet	Created	Number
-----	-----	-----

Negative	12/04/17	2
Negative	12/05/17	12
Negative	12/06/17	35
Negative	12/07/17	13
Negative	12/09/17	22
Negative	12/10/17	31
Negative	12/11/17	24
Neutral	12/04/17	154
Neutral	12/05/17	347
Neutral	12/06/17	875
Neutral	12/07/17	374
Neutral	12/09/17	458
Neutral	12/10/17	730
Neutral	12/11/17	748
Neutral	12/12/17	28
Postive	12/04/17	70
Postive	12/05/17	239
Postive	12/06/17	496
Postive	12/07/17	194
Postive	12/09/17	271
Postive	12/10/17	607
Postive	12/11/17	436
Postive	12/12/17	12

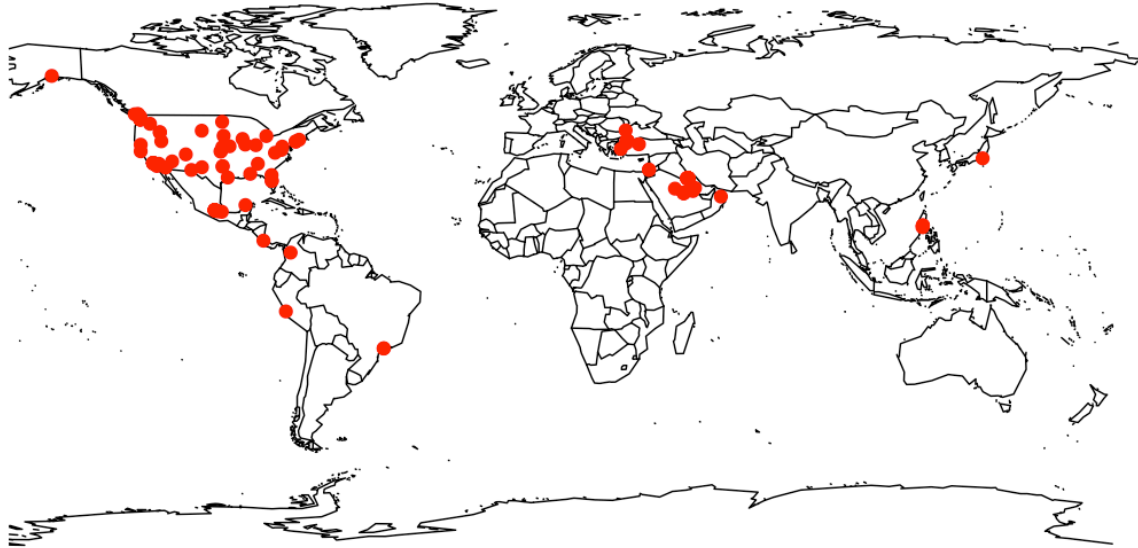
Top hashtags associated with the Brand



This is a ggplot of the top hashtags associated with the Marvel brand. They are mainly all related to the franchise. In the top ten we see "ThorRanganok", "blackpanther" and "Agentsof Shield" (The latest three Marvel movies). With nearly all of the produced hashtags being related to the franchise, we see the extent of Marvel's network. This shows that there are so many other tweets and hashtags being used on twitter that can be traced back to the brand, meaning there are so many users talking about Marvel on twitter. This was calculated by taking the searched and cleaned tweets and extracting other hashtags in the same tweet.

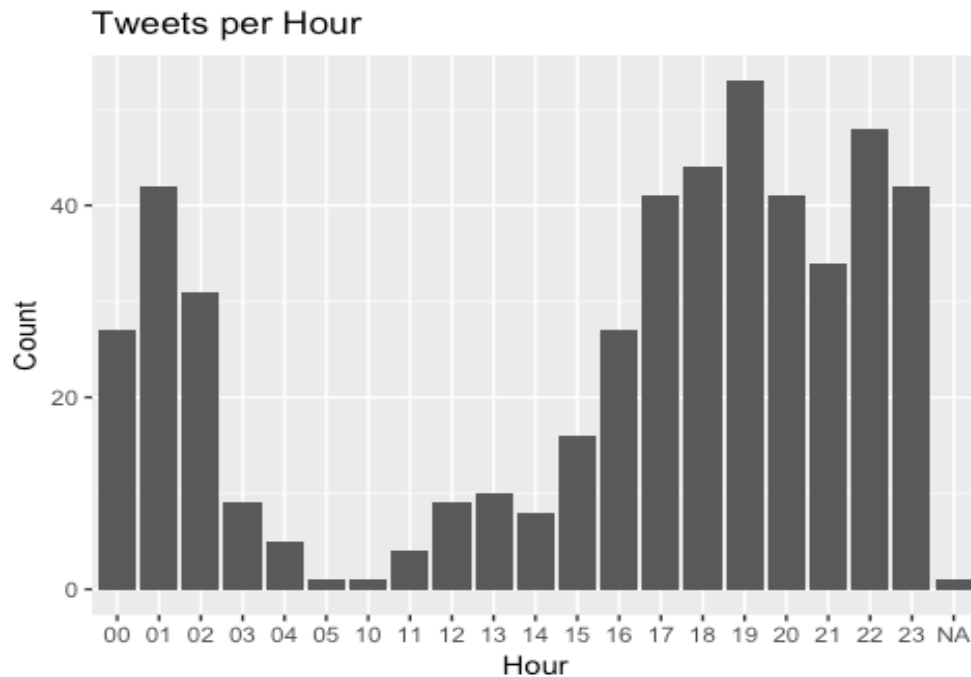
Map (geocoded tweets with keyword "Marvel")

Note: map takes a long-ish time to load since there are not a lot of geocoded tweets! Parsing through about 10000+ tweets



It seems to be that the majority of tweets are originating from the United States, especially from the larger urban and more populated regions and cities. The majority of the other worldwide tweets also seems to be originating from major cities across the globe. We cannot make precise conclusions from the map since so few tweets are geocoded. It does seem fitting however that they are mainly originating from large cities, since they have a higher population. Additionally, it is hard to say if the reason why the majority of the tweets are from the USA is because Marvel's brand is more popular here in America or whether more users in America enable location on their tweets. This was created with ggmap and geocoded tweets (using latitude and longitude).

Tweets per Hour



This histogram shows at what time of the day are people tweeting about marvel! It seems that activity picks up in the evening around 5pm and continues till about 1am. This seems accurate as we can assume that most people would go on social media after their work day or at home in the night. The barplot was created using ggplot and the timestamp on the collected tweets.

Top Tweeters of Hashtags

##	User	Tweets
## 705	Im_a_MARVEL	13
## 718	Iona_Marvel	9
## 1596	TaurusWiggins	7
## 487	DTudo_Com	6
## 108	AMComic	5
## 356	ConradyJW	5
## 642	GtaTv4	5
## 654	hariemackenzie	5
## 532	Erwan_Download	4
## 538	exarmydadof4	4

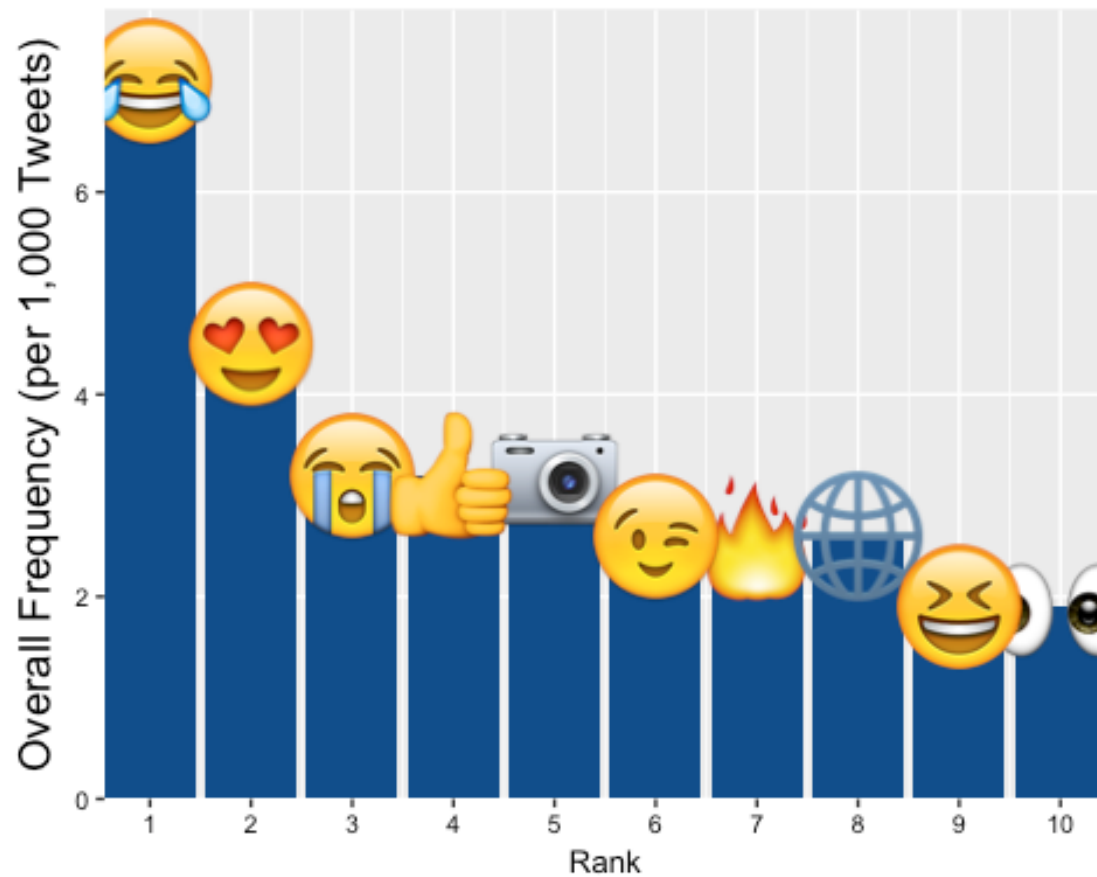
The above table shows the top 10 twitter users that tweet using the hashtag #marvel most frequently. This was created by searching all the collected tweets and seeing if a user has tweeted using #marvel more than once.

Word Cloud



After the initial process of extracting and cleaning the tweets, we use the text mining package (tm) to remove stop words. A stop word is a commonly used word such as “this”. To maintain accuracy, stop words should not be included in the analysis. Next, using the WordCloud package (wordcloud) I create a wordcloud. After "Marvel" (the largest word in the cloud), we have other words relating to the brand, the largest few being "avengers", "jessica", "jones" and “infinity”. In the word cloud, the larger the word, the more often it occurs.

Emoji Analysis (From hamdanazhar)



name	dens	count	rank
camera	3.2	5	5
eyes	1.9	2	10
face with tears of joy	7.1	11	1
fire	2.6	4	7
globe with meridians	2.6	4	8
loudly crying face	3.2	5	3
smiling face with heart-shaped eyes	4.5	7	2
smiling face with open mouth and tightly-closed eyes	1.9	3	9
thumbs up sign	3.2	5	4
Winking face	2.6	4	6