

Movie Review Analysis based on Twitter

615 Final Project Report

GE JIN

12/16/2017

1. Introduction

My primary motivation was to explore reviews of different types of movies posted by users located in different regions based on the data source from Twitter. So I chose five recent movies: **Murder on the Orient Express**, **Coco**, **Justice League**, **Daddy's Home 2**, **Wonder**, as my research subjects.



The analysis mainly includes the exploratory analysis of different types of movies basis their attributes and the sentiment analysis on the tweets posted, to understand customer reviews for these five movies and understand the most frequently appearing words in customer reviews.

The data for this analysis has been collected from twitter from November 24, 2017 to December 8, 2017. This data will be used to visualise the trends basis different attributes and also determine the popular opinions of the people. The text mining approach will be utilized to examine the sentiments of the people basis the twitter texts.

The analysis will help the moviegoers and the movie industry to understand which words customers use to summarize their thoughts about the movie. They can use this analysis to improve box-office results. For the customers, the analysis will help them easily decide what kind of movies to see basis their interests.

2. Exploratory Data Analysis

2.1 Number of tweets

I get 187794 original tweets and 284812 retweets in total. And base on the Table 1, the action movie, Justice League, is most popular, which has 270473 tweets in the 15 days, far more than other four movies. The second one is the animation movie Coco, which has 134471 tweets in total. The least one is the comedy, Daddy's Home 2, only having 8903 relative tweets in those days.

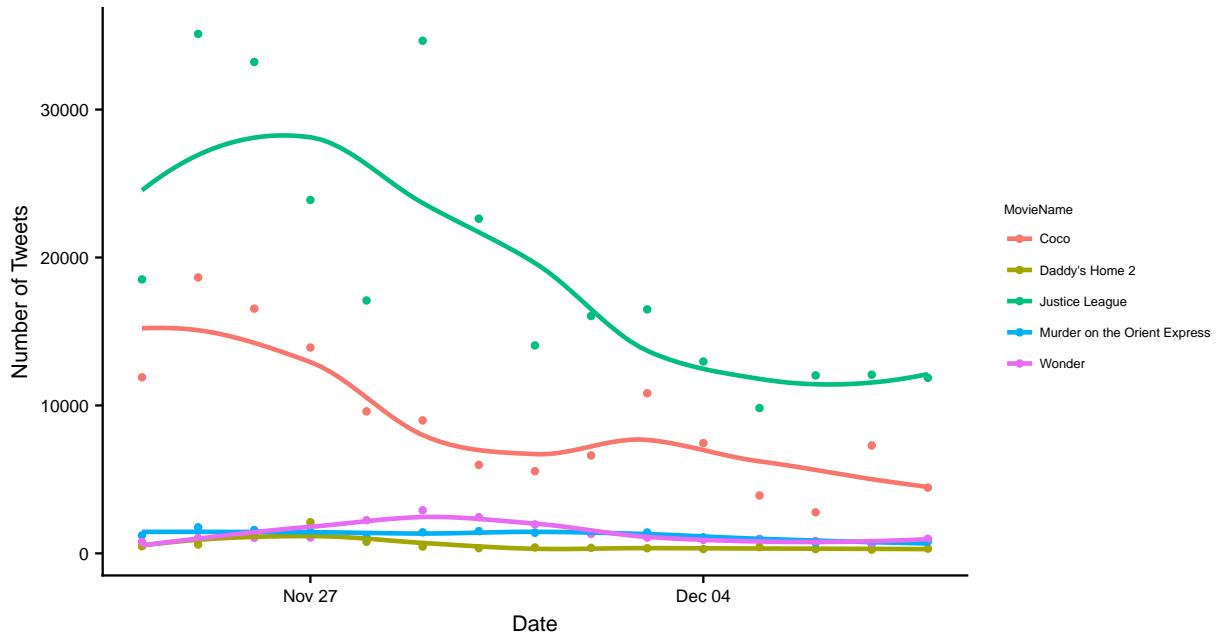
Table 1: Number of original tweets and retweets for different movies

Movie Name	Movie Type	Release Date	Number of Original Tweets	Number of Retweets
Coco	Animation	2017-11-22	35598	98873
Daddy's Home 2	Comedy	2017-11-10	4828	4075
Justice League	Action	2017-11-17	125830	164643
Murder on the Orient Express	Mystery	2017-11-10	13556	4919
Wonder	Drama	2017-11-17	7982	12302

2.2 The Timeline of tweets

From Figure 1, the number of tweets about Murder on the Orient Express, Daddy's Home 2 and Wonder appears to have stabilised in the 15 days, and the number of tweets about Justice League and Coco has shown a declining trend, but they are still much higher than other three movies.

Figure 1: Change of Number of Tweets over Time for Different Movies



2.3 Users

2.3.1 Top Tweeters

Figure 2-6 show the top 20 users' screen names of twitter for different kinds of movies. To be more specific, user thePositiveMOM_ tweeted more than 600 times about Coco, and user Saequann posted more than 400 tweets about Justice League. User greg_ivan is much more interested in Murder on the Orient Express than other users. And user swindonguide and auk_sanejourno are interested in Daddy's Home 2 and Wonder respectively.

Figure 2: Tweets about Orient Express Movie by Top 20 Users

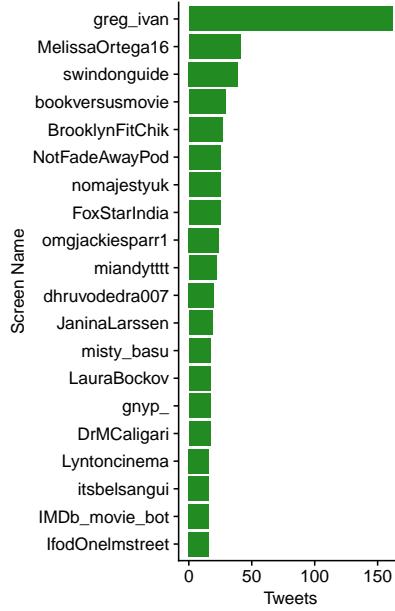


Figure 3: Tweets about Coco by Top 20 Users

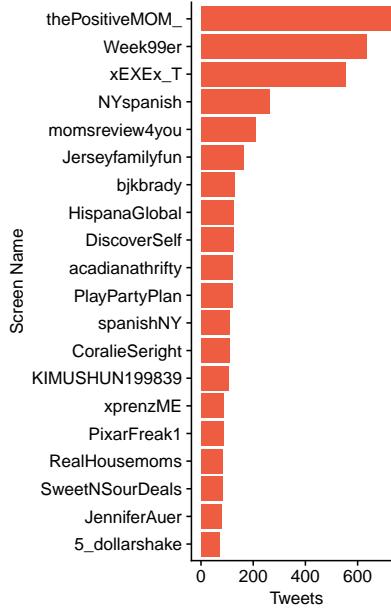


Figure 4: Tweets about Justice I by Top 20 Users

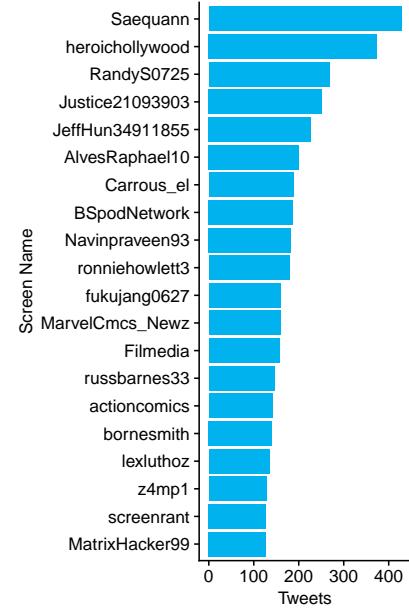


Figure 5: Tweets about Daddy's Home by Top 20 Users

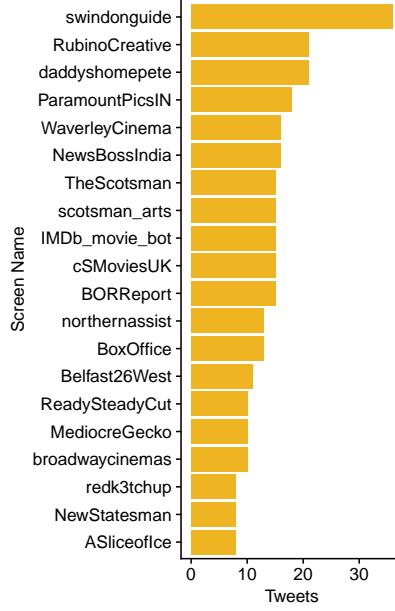
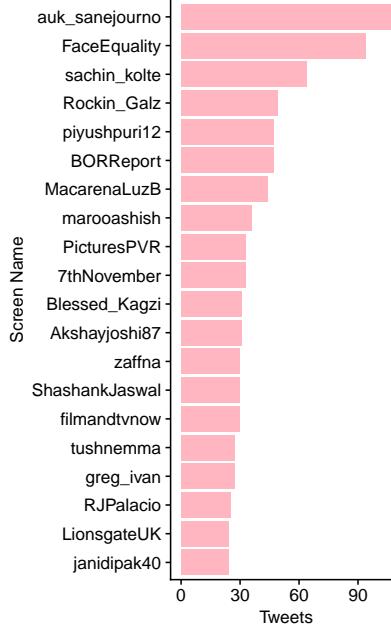


Figure 6: Tweets about Wonder by Top 20 Users



2.3.2 Platforms of Users

From Figure 7, above 70% tweets are posted by twitter webs, iphones and Andoid mobile phones, which are mostly what we would expect, but some bots got in on the action too. And other common platforms used in twitter are Twitter Lite, Twitter for iPad, TweetDeck.

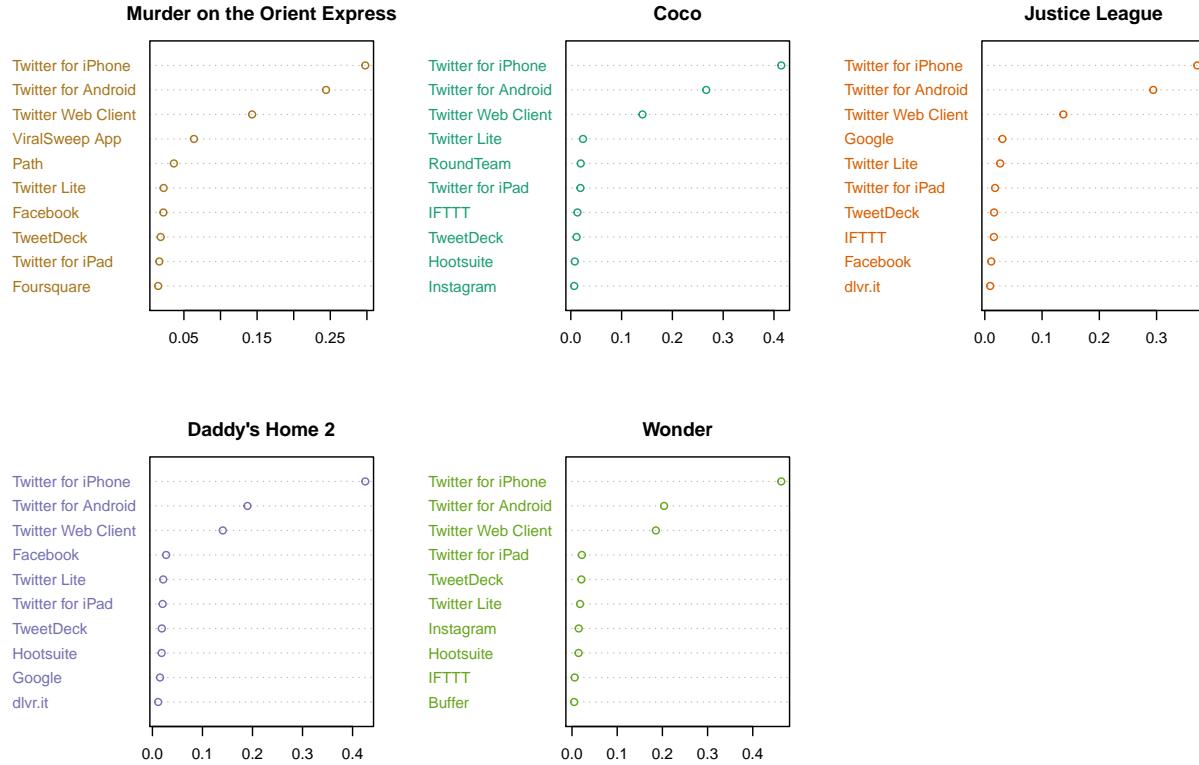


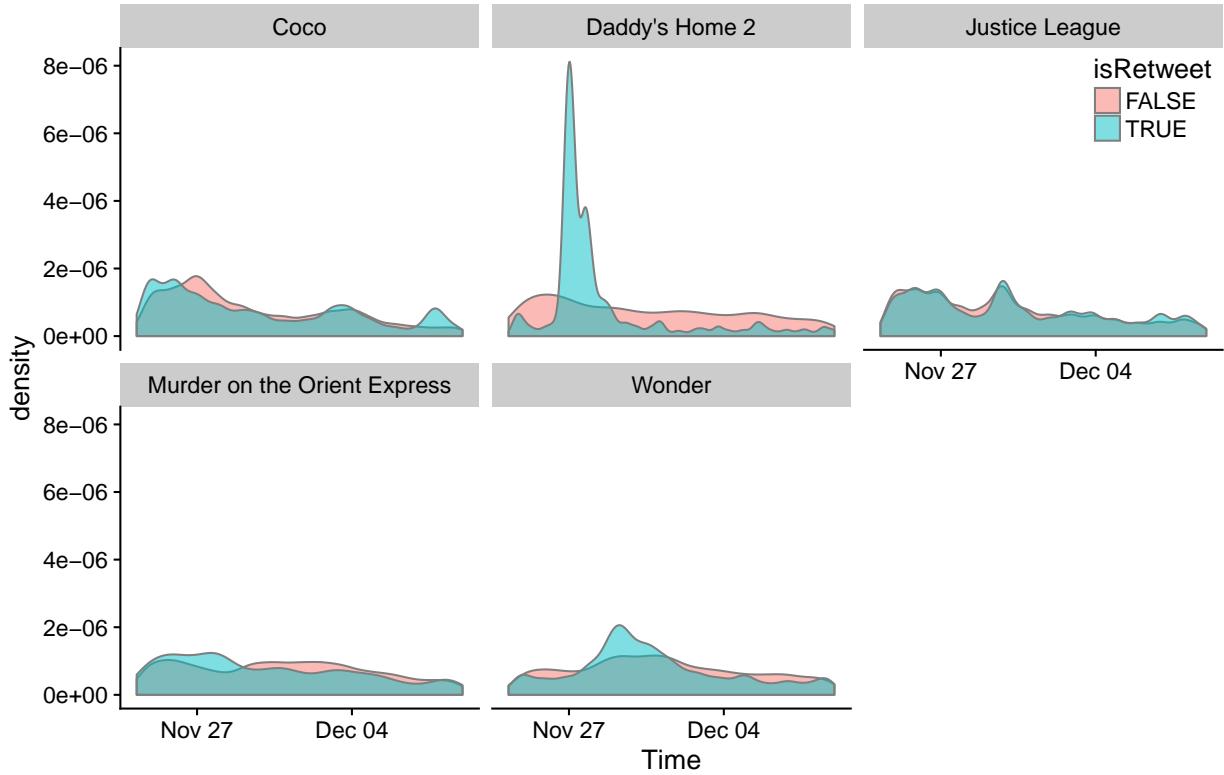
Figure 7: Percentage of Tweets about Different Movies posted by Different Platforms

2.4 Retweets

2.4.1 Timeline of Density of Original Tweets and Retweets

In the 15 days, all the movies have similar variation trend of density of original tweets and retweets, except for Daddy's Home 2. The density of retweets about this movie is extremely high around Nov 27 and Nov 28 these two days.

Figure 8: Density of Original Tweets and Retweets over Time for Different Movies



2.4.2 Retweet Network

The ‘retweet network’, composed from the original author of the tweet to the retweeter. And we can get a sense of who is being retweeted. We can see a core of individuals engaging in a conversation at the center of the following graphs. Nodes are sized to their total degree (retweeted and being retweeted), and edge-width is proportional to the number of retweets between that pair. Labeled nodes are those that were retweeted at least once.

Figure 9: Retweet Network of Orient Express Movie

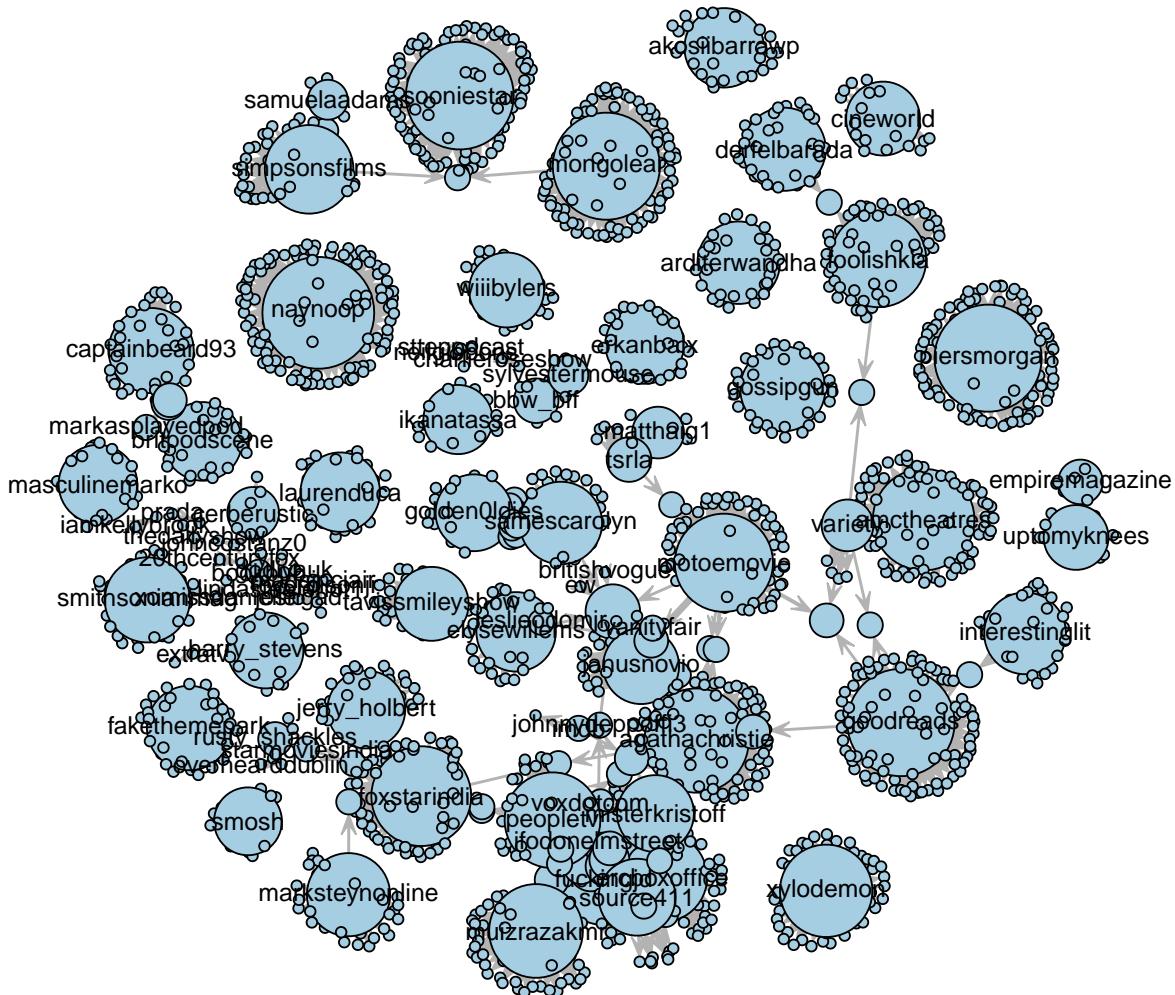


Figure 10: Retweet Network of Coco

disney pixar

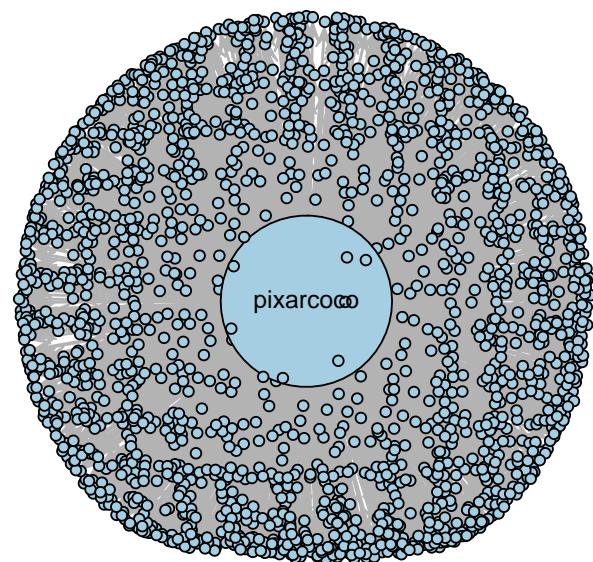


Figure 11: Retweet Network of Justice League

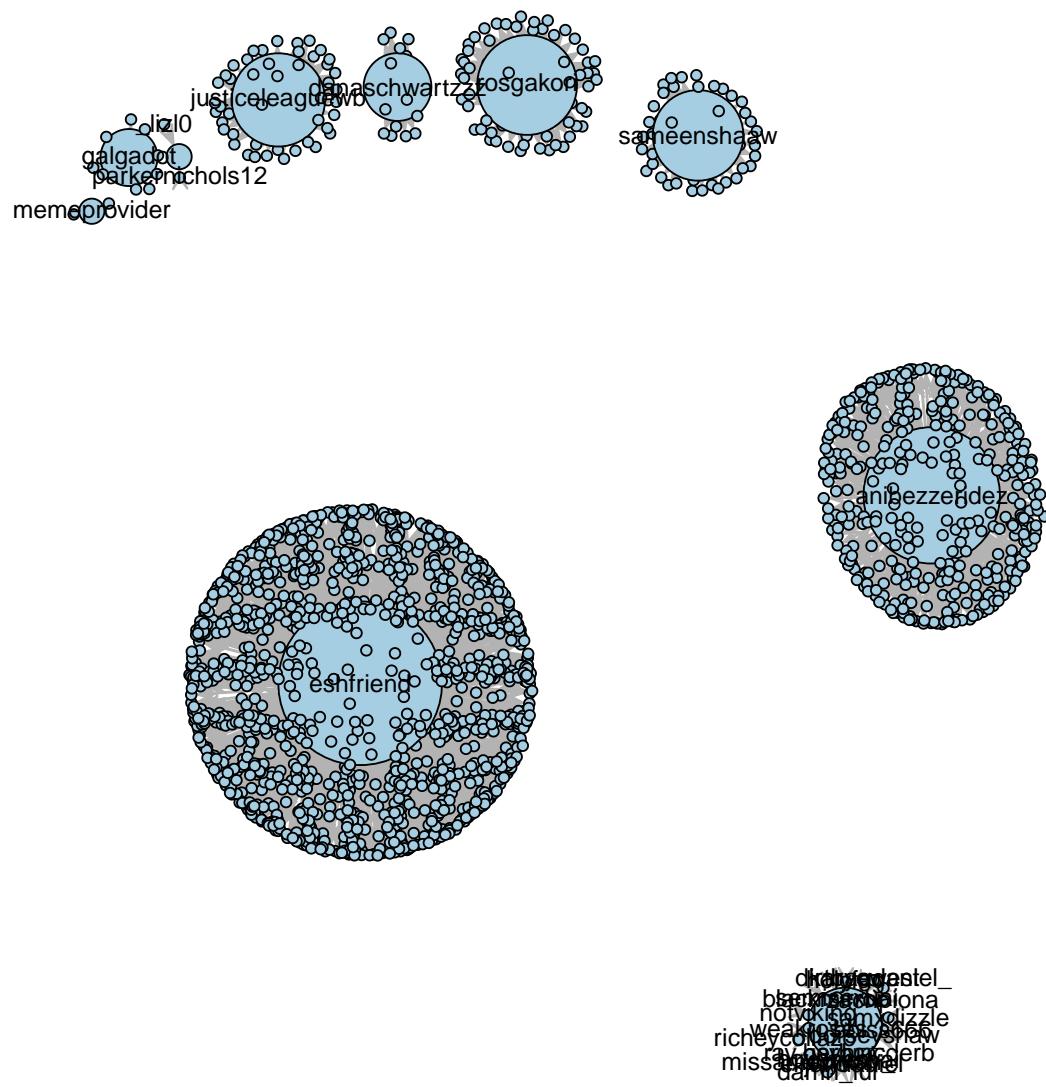


Figure 12: Retweet Network of Daddy's Home 2

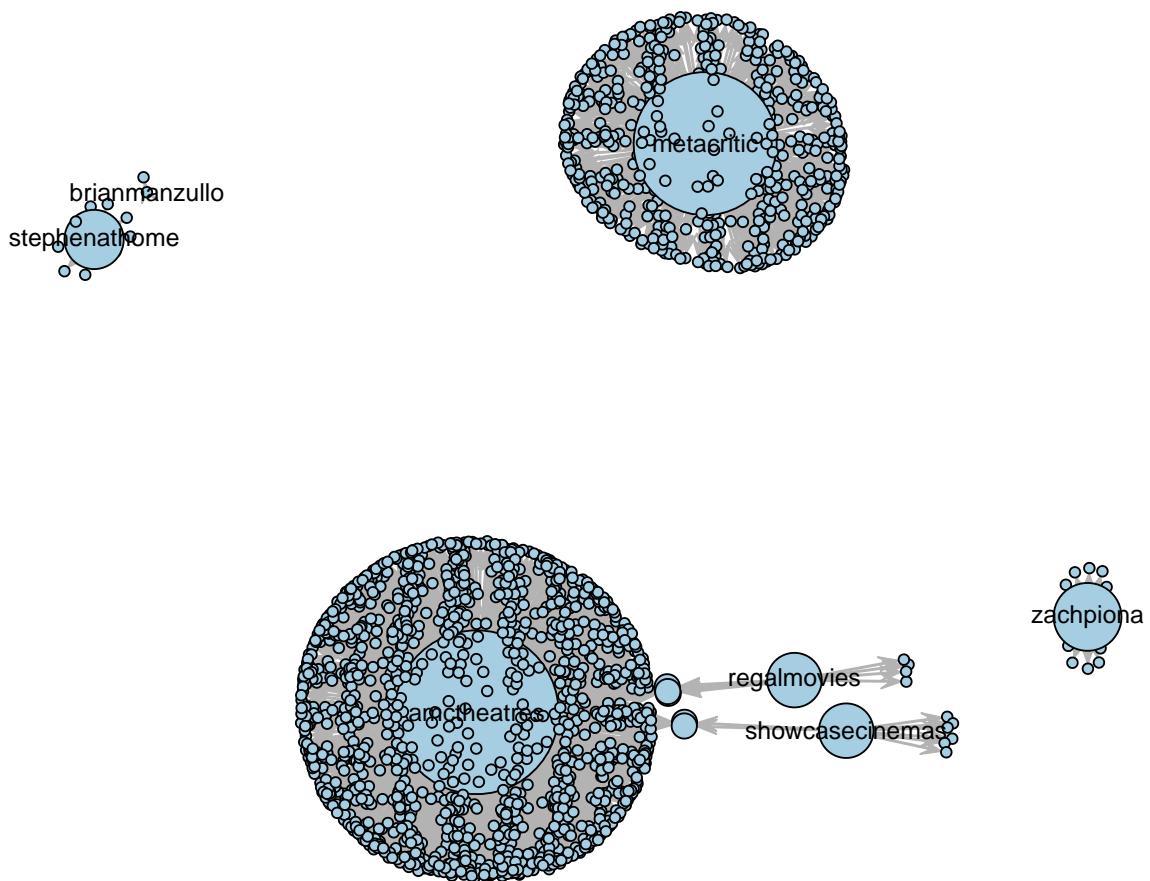
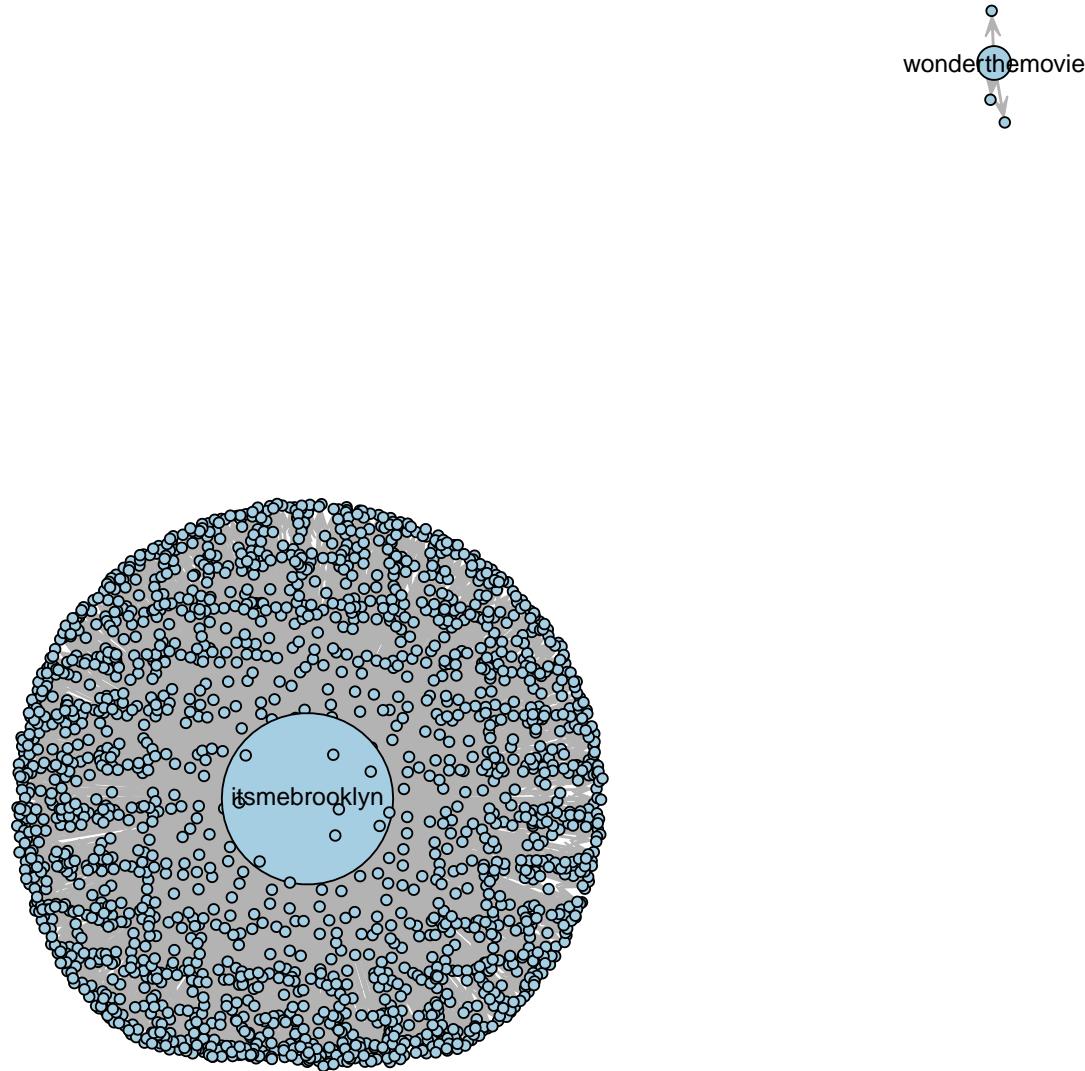


Figure 13: Retweet Network of Wonder



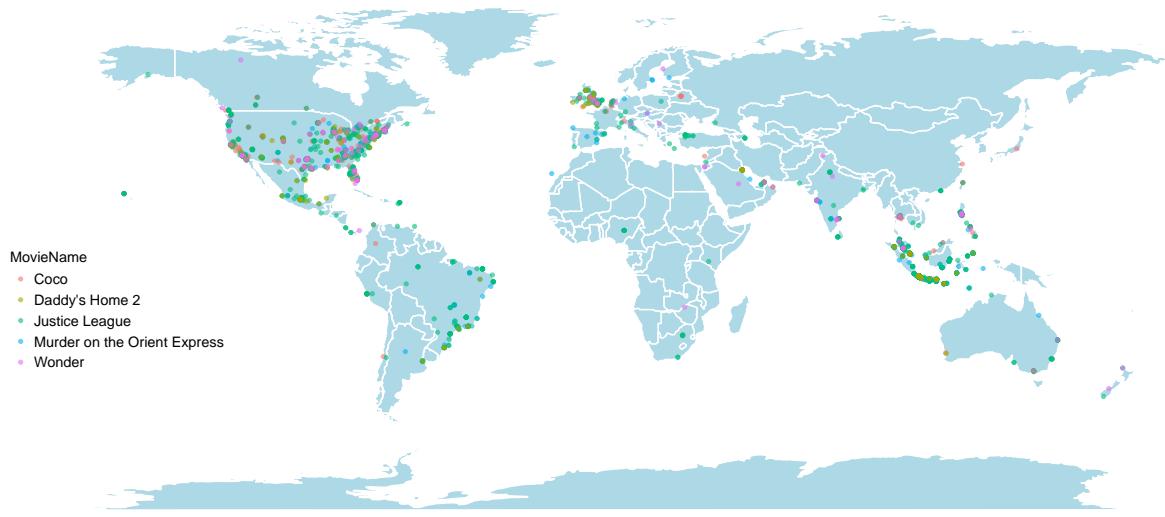
2.5 Geolocation of Tweets

Although you might be surprised with the small number of tweets on the map, typically only 1% of tweets are geocoded. I collected a total of 472606 tweets, but only 2327 (around 0.502% of total tweets) were geocoded.

2.5.1 World Map

From the world map, we can see that most tweets were geocoded in the region of United States. So the following analysis is focus on the USA region.

Figure 14: World Map of Tweets

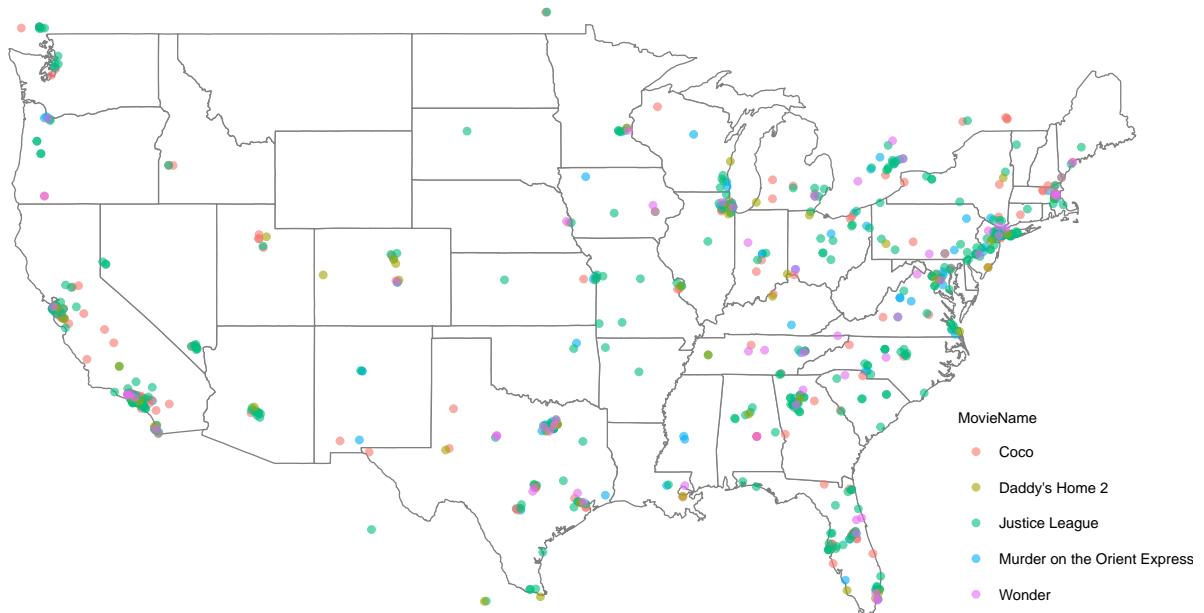


2.5.2 USA Map

Only 863 geocoded tweets are located in USA.

From the USA map, does not include the two states, Alaska and Hawaii, tweets are mainly geocoded in Eastern and Western America, with few points showing in Middle America. Besides, it is clear that most tweets are about Justice League.

Figure 15: USA Map of Tweets



2.5.3 States Map

Next, I focus on exploring the difference between states in America about the number of tweets of different movies. From Figure 16-20, the tweets were mainly located in the following states: California, Texas, Florida and New York.

Figure 16: Density of Tweets about Orient Express Movie in Different

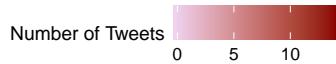
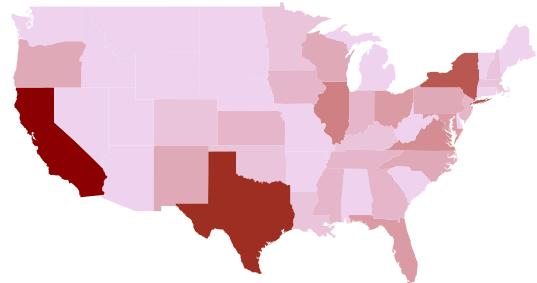


Figure 17: Density of Tweets about Coco in Different States in USA

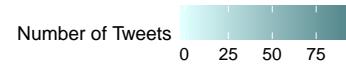
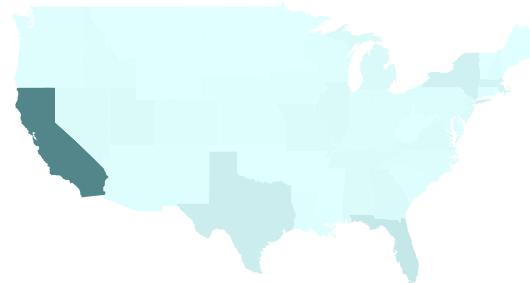


Figure 18: Density of Tweets about Justice League in Different States

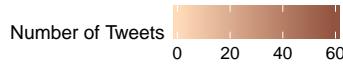
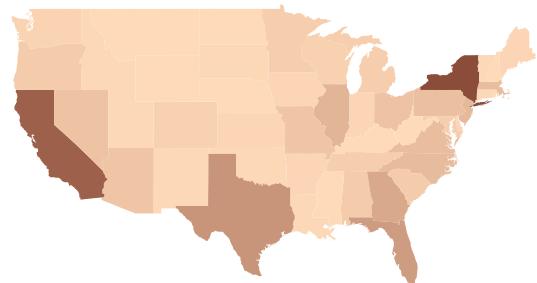


Figure 19: Density of Tweets about Daddy's Home 2 in Different State

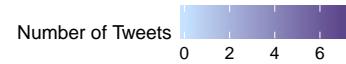
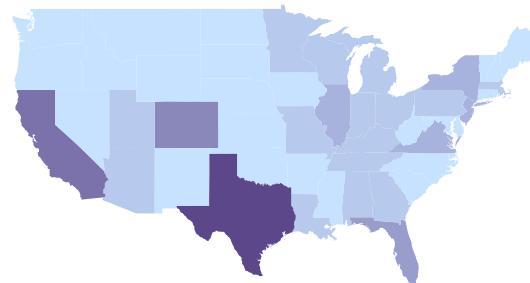
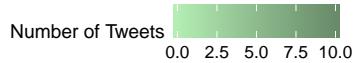
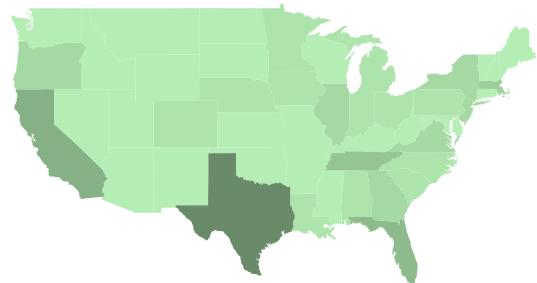


Figure 20: Density of Tweets about Wonder in Different States in USA



2.6 Trending Words across Movie Reviews

This section here analyses the words that were trending in the reviews of five movies. This will help to determine how these words contribute towards the evaluation of the movies and also, how different people react to the movies.



2.7 Sentiment Analysis

In this section, we first show the sentiments associated with most common words. Then to determine the sentiment score associated with the word, the ‘nrc’ lexicon in the sentiment dataset, available in the tidytext package, has been used.

2.7.1 Most Common Positive and Negative Words

This section shows the most common words appearing in the tweet text with sentiment for different types of movies. The word ‘win’ is shown repeatedly in those movies: Coco, Justice League, Daddy’s Home 2 and

Wonder; and in Murder on the orient express, the word ‘murder’ shows most commonly.

Table 2: Common Words about Orient Express

Word	Sentiment	Count
murder	negative	16441
win	positive	1224
mystery	negative	481
love	positive	236
death	negative	215
amazing	positive	201
top	positive	186
enjoyed	positive	175
pretty	positive	169
classic	positive	165

Table 3: Common Words about Coco

Word	Sentiment	Count
win	positive	21168
prize	positive	14432
frozen	negative	7821
perfect	positive	3546
love	positive	3045
beautiful	positive	2923
critics	negative	2742
cry	negative	2236
amazing	positive	2101
challenging	negative	2047

Table 4: Common Words about Justice League

Word	Sentiment	Count
win	positive	7072
love	positive	5804
marvel	positive	5324
parody	negative	4971
injustice	negative	4308
effectively	positive	3681
bad	negative	3247
cool	positive	3038
prize	positive	2962
respect	positive	2923

Table 5: Common Words about Daddy’s Home 2

Word	Sentiment	Count
win	positive	1518
prize	positive	1450
funny	negative	276
bad	negative	205
hilarious	positive	173
top	positive	121
murder	negative	114
fucking	negative	61
love	positive	57
recommend	positive	56

Table 6: Common Words about Wonder

Word	Sentiment	Count
win	positive	496
magic	positive	380
love	positive	326
amazing	positive	313
kindness	positive	307
loved	positive	282
beautiful	positive	265
hug	positive	251
wonderful	positive	251
perfect	positive	249

2.7.2 Time trend of Sentiment

The nrc lexicon categorizes words in a binary fashion ('1 or 0') into categories of positive, negative, anger, anticipation, disgust, fear, joy, sadness, surprise, and trust. Then, the average of such scores for each day has been calculated and information about how people think about this movie is retrieved. I depicted the information using line graphs as shown below.

Figure 21: Sentiment Over Time about Orient Express

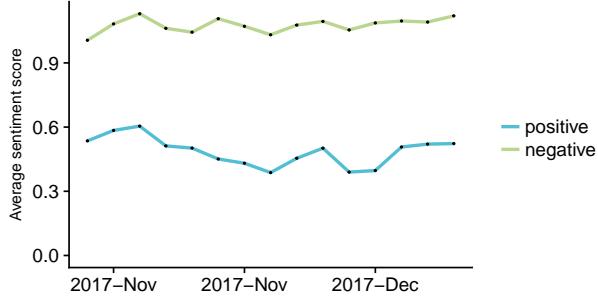


Figure 22: Sentiment Over Time about Coco

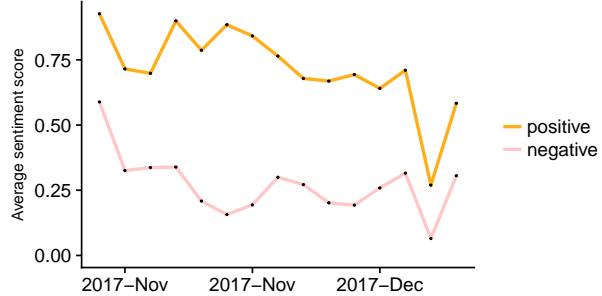


Figure 23: Sentiment Over Time about Justice League

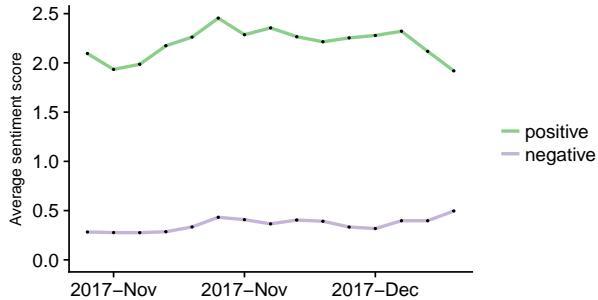


Figure 24: Sentiment Over Time about Daddy's Home 2

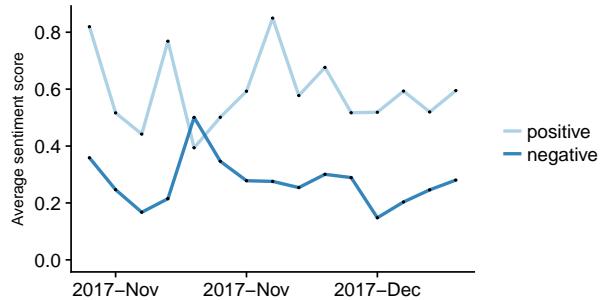
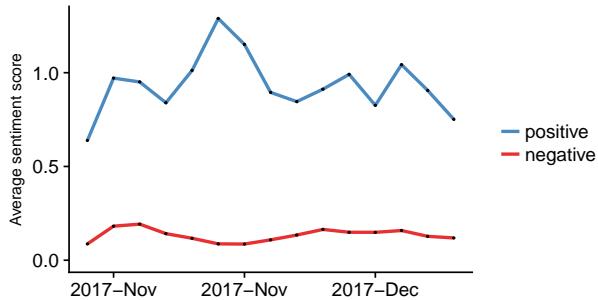


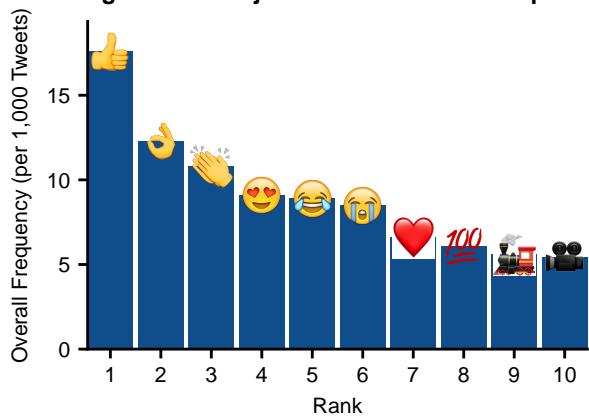
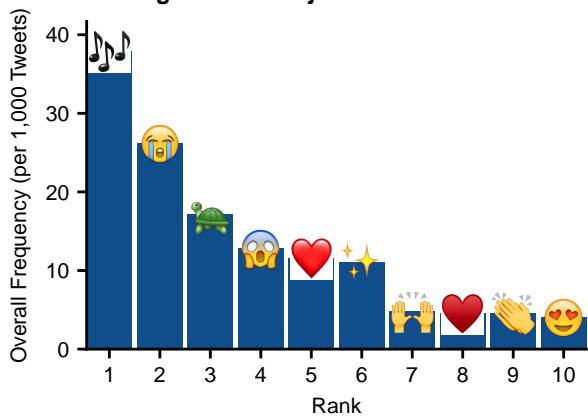
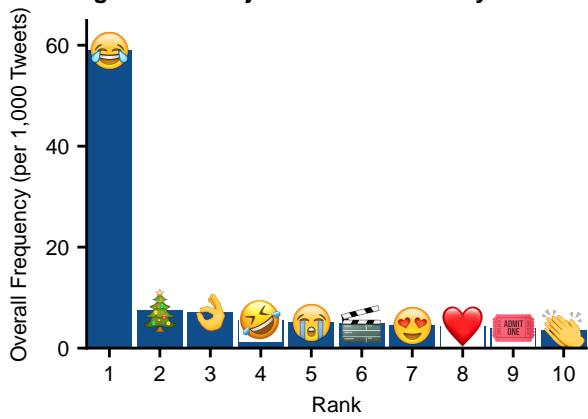
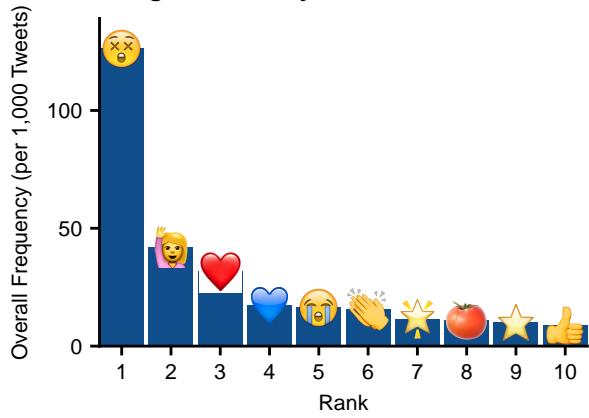
Figure 25: Sentiment Over Time about Wonder



As a result, the sentiment of all these movies tends to positive, except for Murder on the Orient Express.

2.8 Emojis Analysis

The following shows the most used emojis in different movies, to explore how people react to the movies. Since the data about Coco and Justice League are too large, the data of these two movies used for analysing emoji reactions is selected by random sampling of size 20000. From the graphs shown below, we can get a sense of how people react to those movies.

Figure 26: Emoji Reactions to Orient Express**Figure 27: Emoji Reactions to Coco****Figure 28: Emoji Reactions to Justice League****Figure 29: Emoji Reactions to Daddy's Home 2****Figure 30: Emoji Reactions to Wonder**

3. Conclusion

With the use of plethora of features in R, the analysis will provide key insights about those five movies, including timeline of tweets, user platforms of tweets, retweet network, etc. Also, maps were made to locate the tweets. The data was then reviewed graphically to explore the trending words in twitter about those movies and sentiment analysis was done using word cloud, line plots and horizontal bar charts, to tell us that

how different customers react to different types of movies.

Various results and analysis showed that most people in USA prefer action movies, then animation movie, like Coco. The comedy Daddy's Home 2 is least noticed in twitter. Besides, based on the states maps, people in California, Texas, Florida, New York and Virginia tweets more about those movies. But sentiment analysis, and emoji analysis, all together show us positive reactions of customers associated with the movies. win, love, prize, and Perfect - customers associate movie with positive words like these. This can help movie industry in understanding how branding and associating with these words can further help them improving box office.