All the Feels:

Sentiment Analysis

Between Emoji and Text

Alexandra Plassaras

Master’s Candidate, 2017

Columbia University

Quantitative Methods in Social Science

Master’s Thesis

Table of Contents

Introduction 3

Research Problem 4

Hypotheses 4

Previous Research 6

Research Design 8

Data 8

Sentiment Analysis - Text 11

Sentiment Analysis - Emoji 12

Data Processing 13

Results 15

Discussion 25

Results 25

Limitations 27

Conclusion 32

Appendices 34

Appendix A – QMSS G4063 Information 34

Appendix B – All variables associated with a single tweet 34

Appendix C – Unicode List of Emoji 34

Appendix D – Emoji Sentiment Lexicon 35

Appendix E – Additional Visualizations 36

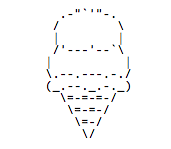
References 38

# Introduction

Macintosh HD:Users:alexandraplassaras:Desktop:Screen Shot 2016-12-02 at 9.42.49 PM.pngandMacintosh HD:Users:alexandraplassaras:Desktop:Screen Shot 2016-12-02 at 9.42.57 PM.png. These are just two novels that have in the past few years been translated into emoji. Can you guess which classic novels these emojis are referring to? If you guessed Alice in Wonderland[[1]](#footnote-1) and Moby Dick[[2]](#footnote-2) then you are correct. In recent years we have seen a new way to express ourselves in online and mobile communication. As of 2015, emojis have become the world’s fastest growing language in all forms of communications – social media, text messaging and various messaging apps and even email.[[3]](#footnote-3) A survey conducted by TalkTalk Mobile, a British mobile retailer, found that 72% of 18 to 25 year olds stated that emojis were easier to use to express their feelings than text[[4]](#footnote-4). Knapp and Hall claimed that emojis serve as nonverbal conversational cues “help to communicate ideas, manage interactions and disambiguate meaning to improve the efficiency of the conversation”.[[5]](#footnote-5)

Originally created for use in Japanese mobiles in the late 1990s, support for emoji became available to major mobile operating systems iOs and Android in the US in 2011.[[6]](#footnote-6) As of 2016, research had suggested that emojis have already taken over emoticons on social media most likely due to their flexibility in expressing not only facial expressions but food, religion, activities and even various cultures.[[7]](#footnote-7)

Prior to the widespread use of emojis, emoticons were widely used to express feelings, moods and emotions. An emoticon is shorthand for a facial expression – such as : - ) or : - ( . Emojis are emoticons on steroids – instead of using alphanumeric, punctuation and logic symbols[[8]](#footnote-8), emojis are graphic symbols that represent facial expressions as well as concepts and ideas.[[9]](#footnote-9) Prior to the introduction of emojis it was not possible to conveyor  using emoticons without significant time and creativity as seen in Figure 1 below.



**Figure 1** – Ice cream emoticon

Now however, actions, religions, cultures, animals and plants can all be expressed using emojis. Given the prevalence of emoji in our daily communication, the focus of this research is to explore the sentiments that emojis attempt to convey using sentiment analysis of Twitter data and how this compares to the tweet’s text.

# Research Problem

## Previous Research

Much research has been conducted on text sentiment analysis ranging from subjectivity and sentiment analysis[[10]](#footnote-10) to detecting sarcasm in sentiment analysis.[[11]](#footnote-11) As the focus of this study is emoji sentiment analysis and how it compares to text sentiment analysis, the majority of this section will focus on previous research done concerning understanding emoji. Given that emojis were introduced to Americans in 2011 on a large scale when Apple added the emoji keyboard to its products, existing studies on emoji are quite limited. Additionally as Lu *et al.* mentions, it is harder to come across large data sets of emoji usage.[[12]](#footnote-12)

The research uncovered so far includes a global analysis of emoji used on smartphones via the Kika Emoji keyboard, one of the most popular third party keyboards on Android smartphones.[[13]](#footnote-13) This study looked at over 400 million emoji-contained messages from users in 212 different countries and showed that there is a difference in emoji usage based on country and region. Another study looked at how people interpret different emojis as well as the same emojis on different platforms.[[14]](#footnote-14) Subjects were surveyed from various countries including the US, the UK, Canada, Brazil and Germany and were recruited from social media. This study made the claim that people do in fact interpret emojis differently on an individual basis and not just from a cultural and country basis.

Previous work done by Novak *et al.* significantly affects the scope of this project because their study created the first known open-source emoji sentiment lexicon, referred to as the Emoji Sentiment Ranking.[[15]](#footnote-15) This study classified over 1.6 million tweets in 13 different European languages (including English) to create a sentiment strength measure lexicon for every emoji. Of roughly 64,000 tweets only 4% of the tweets contained emojis. What they found was that the majority of emojis were positive and that the sentiment of tweets with and without emojis varied greatly. The emoji lexicon created through this work will be used as the emoji lexicon of this study.

Previous research has not focused on comparing sentiment of text with emoji and instead has typically been either text or emoji. Furthermore, the research on emoji has been conducted on either a small scale (under 30,000 instances) or using platform specific data from mobile carriers or specific applications. There have not been studies that have looked at sentiment analysis of emojis in Tweets. Thus, this paper hopes to add insights on the differences between text and emoji sentiment found in Tweets. Since emoji is a popular and effective form of communication, textual analysis that disregards emoji use would be insufficient or incomplete. Removal of emoji characters in NLP analysis may result in less accurate analysis of the text being studied, thus this paper aims to show the importance of emojis in sentiment analysis.

## Hypotheses

This study will focus on three hypotheses. The first hypothesis is that there will be a significant difference between sentiment analysis of text compared to the sentiment analysis of emojis. This is because this paper believes that there will be more frequent sarcasm used on this social media platform. For example, while the text “Make America Great Again” may be coded as having positive sentiment, this paper believes that the emojis will not always share the same sentiment and will perhaps be negative (e.g. using sad, crying or angry faces). This paper is also assumes that there are more urban liberal Twitter users than there are rural conservative users and thus tweeting sarcastic text and emojis would be used more likely among urban liberal Twitter users. Boia *et al.* looked at emoticons and their relationship with text sentiment and came to the conclusion “that the sentiment conveyed by an emoticon generally agrees with the sentiment of the entire Tweet.”[[16]](#footnote-16) However this paper argues that given the vast number of emojis that exist, many of which are not faces and instead food emojis, inanimate objects and shapes that the relationship between emojis and text will be different than the relationship between emoticons and text.

The second hypothesis is that tweets referencing Republican candidates (e.g. Trump, Cruz and Rubio) will contain more negative emojis than tweets referencing Democratic candidates (e.g. Clinton and Sanders). In the example mentioned above this paper hypothesizes that people discussing Republican candidates who use emojis are more likely to be young users who might be are more liberal and more likely to use emojis in a sarcastic and negative manner.

The third hypothesis is that tweets that mention Democratic candidates are more likely to have higher negative text sentiment than tweets mentioning Republican candidates. This paper hypothesizes that Twitter users who reference a candidate are more likely to use harsher and more negative language if they are referring to a Democratic candidate. This may be because of 1) some Democrats’ strong affiliation with Sanders and the ‘Never Clinton’ Campaign[[17]](#footnote-17) or 2) Republican voters who were either against Sanders or Clinton. Because of the recent rise in hate crimes[[18]](#footnote-18) after the popular vote that named Donald Trump President-Elect this paper assumes that people who were more likely to support Trump were more likely to use vocabulary similar to Trump’s rhetoric: “ ‘losers,’ ‘total losers,’ ‘haters,’ ‘dumb,’ ‘idiots,’ ‘morons,’ ‘stupid,’ ‘dummy’ and ‘[disgusting](http://lastnighton.com/2015/08/02/jimmy-kimmel-spotlights-donald-trumps-love-of-the-word-disgusting/)’ ”[[19]](#footnote-19) particularly when it came to referring to the opposite political party and it’s candidates.

# Research Design

## Data

The data used for this project consists of Tweets with specific hashtags collected from February 7, 2016 to April 2, 2016 that referred to the 2016 presidential primaries. Approximately 1.37 GB of tweets were collected during this time period. Of the tweets that were collected, 1,816,475 (approximately 318.4 MB) were captured that were in English and contained geo-location data with 237,499 unique users. The geo-location attribute was used to filter out all tweets that were not identified as sent from the United States. As this study is looking at American sentiment in both text and emoji making this assumption ensures a higher likelihood that the tweets we will be looking at are coming from Americans. Thus tweets that do not include geo-location information are not used in the following study as there is no way to identify where in the world they were sent from.

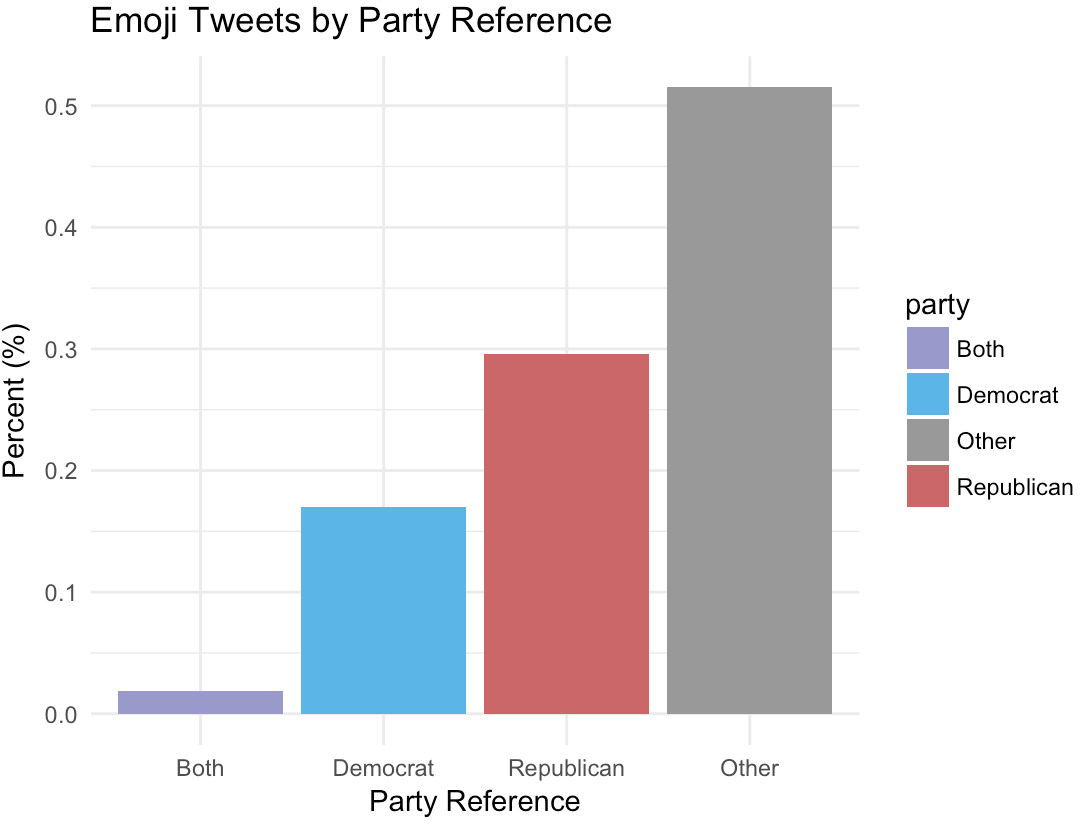
The tweets used in this study were scraped via Twitter’s Streaming API during a 56 day period and information regarding both meta data about the tweets as well as the tweet and user data were collected. Of the initial 1,816,475 Tweets in the data set, only 86,258 (~4.75%) contained emojis. A total of 50 variables associated for each tweet have been collected but this research will only look at a small subset of these variables. For the full list of variables collected for each tweet instance refer to Appendix B.

The variables used in this study are unique identifiers for each tweet, usernames, tweet contents and locations of each tweet. The tweets were selected from Twitter’s API based on their reference to one of any of the following candidates – Hillary Clinton, Bernie Sanders, Ted Cruz, Donald Trump, and Marco Rubio. Nicknames and references to particular candidates were also included such as Trumpf, Hillary and Cruz. Figure 2 below shows the full list of identifiers used to pull tweets from Twitter’s API. For more information on the data collected, see Appendix A.

|  |  |
| --- | --- |
| **Candidate’s Full Name** | **Key words associated to locate Tweets** |
| Hillary Clinton | Clinton, clinton, Hillary, hillary, Hillaryclinton, hillaryclinton, Hillary Clinton, hillary Clinton, HillaryClinton |
| Bernie Sanders | Berniesanders, berniesanders, Bernie Sanders, bernie sanders, Bernie, Bernie, Sensanders, SenSanders |
| Ted Cruz | Cruz, cruz, Ted, ted, Tedcruz, tedcruz, Ted Cruz, ted cruz |
| Donald Trump | Donaldtrump, donaldtrump, Donald Trump, donald trump, Trump, trump, Donald, Donald, Trumpf, trumpf, realDonaldTrump |
| Marco Rubio | Marcorubio, marcorubio, Marco Rubio, marco rubio |

**Figure 2 –** Breakdown of Identifiers for each Candidate

Furthermore only 41,834 tweets (~2.30% of the original data set) contained references to the Democratic or Republican party. As two out of the three hypotheses for this analysis deal with sentiment across political parties, the following results are based on the 41,834 tweets that contain emojis and references to a political party. Figure 3 below shows the distribution of tweets as categorized into political parties. The largest group of tweets are those classified as “other” at 51.5% meaning that they were not classified into any political party. The second largest group was the Republican party at 29.6% whereas the Democratic party consisted of 17% of tweets. A final fourth category contained tweets referencing both Democrat and Republican parties at 1.9%. The other category being the largest category is not particularly surprising given the limitations of subsetting tweets using key words like the primary candidate names and the parties themselves because there may be other references to a particular party that are more nuanced and difficult for an algorithm to decipher. What is surprising is that there are more tweets with emojis referencing the Republican party than there are the Democratic party. This may be because of the use of sarcasm towards Republicans or it might mean that people who Tweet about the Republican party might be more emoji-literate.



**Figure 3** – Emoji Tweets by Party Reference

To identify the emojis currently in the dataset a full list of all available emoji will be used. This full list comes from the Unicode Consortium which was scraped from their website. More info on the Unicode Consortium can be found in Appendix C. The reason that this list of emojis will be used is because this list is a complete list of all globally recognized emojis, which Twitter also uses. Choosing this globally recognized list guarantees that all emojis in the data set will be accounted for because users are unable to input emojis that are not recognized by the Unicode Consortium. There are a total of 2,389 recognized emojis within this list. As described below in further detail in the sentiment analysis section, the lexicon for emoji sentiment is not as complete as this full list so there may be emojis in the data set that do not have a sentiment score.

Lastly, once sentiment was given to both the text and the emoji subsets of tweets, only 26,026 tweets were identified as having sentiment. This constraint was primarily due to the text subset as the lexicon used in this study was only able to identify this amount of tweets as having some sort of sentiment. Within this subset which will be looked at throughout the rest of this study, there were 12,817 uniquely identified users who tweeted a total of 26,026 tweets with text and emojis present that contained identifiable sentiment strength and referred to a political party. Within these tweets, 1,453 unique words and 658 unique emojis were identified as containing some value of sentiment strength.

## Sentiment Analysis - Text

As defined by Taboada *et al.* sentiment analysis refers to a method of extracting subjectivity and polarity from text.[[20]](#footnote-20) The polarity of the text is on a scale of positivity, neutrality or negativity. Mathematically represented by:

Two main methods of sentiment analysis exist – the lexicon-based approach and the text classification approach.[[21]](#footnote-21) This study utilizes the lexicon-based approach and calculates the orientation of text from the semantic orientation of the words in the tweet.[[22]](#footnote-22) To do this, the lexicon developed by Finn Årup Nielsen called Afinn which contains 2477 words will be used as a basis on which to calculate text sentiment.[[23]](#footnote-23) With this lexicon words were given a value ranging between . These values were then reduced to by dividing every sentiment score by 5 so that the range is the same range that the emoji lexicon follows.

Another popular lexicon that was considered in this study was Liu and Hu’s lexicon which contained 6,859 words. This lexicon focused on positive/negative polarity and while the lexicon itself contained more words than the Afinn lexicon it identified only 24,457 unique words from the tweet corpus compared to the Afinn’s 26,026 unique words. Thus Liu’s lexicon was not used in this study. Instead the Afinn lexicon was chosen as it focused on sentiment strength instead of positive/negative polarity. Additionally the Afinn lexicon was designed for sentiment analysis of microblogging media platforms such as Twitter and is therefore a more appropriate lexicon for use in this study. As a result of focusing in microblogging sites the lexicon has added sentiment strength scores to common phrases found on the internet such as “lol” (laughing out loud) as well as strong curse words into it’s lexicon.[[24]](#footnote-24)

The Afinn lexicon was originally created in 2009 to look at sentiment analysis of tweets focused on the United Nation Climate Conference (COP15).[[25]](#footnote-25) The first version of the lexicon was referred to as AFINN-96, which contained 1,486 unique words with a small amount of phrases. A second version was released in 2011 that contained 2,477 unique words as well as additional phrases not included in the original release.[[26]](#footnote-26) Some notable additions to the lexicon that were not included in other lexicons include “lol”, “wtf” and “rofl”. The sentiment strength for each word in the Afinn lexicon was calculated by taking the sum of the word’ sentiment strength divided by the number of words represented.

## Sentiment Analysis - Emoji

The emoji lexicon used in this study was developed by Novak *et al.* andcontains 751 unique emojis that occurred at least 5 times in their data set of 70,000 tweets with emojis.[[27]](#footnote-27) These emojis were cross-referenced with emojitracker, a website that monitors in real-time the use of emojis on Twitter, in June of 2015.[[28]](#footnote-28) The study then looked at the Pearson correlation for emojis with N >= 5 and determined that they were highly significant at the 1% level, confirming that the list of emojis chosen for this lexicon was representative of the emoji’s general use on Twitter’s platform. More information on the emoji lexicon can be found in Appendix D.

## Data Processing

In order to conduct sentiment analysis on both the text and emoji in the data set the following steps were taken. Re-tweets and repetitive tweets were removed from the data set. If not removed the likelihood that they would skew the results would have increased. In order to determine the sentiment for each tweet’s emojis a dictionary was created which identified each emoji character to it’s unique identifier and it’s name. Multiple open source emoji dictionaries such as twitterEmojiProject[[29]](#footnote-29), Prismoji’s emoji tutorial[[30]](#footnote-30) and Twimoji[[31]](#footnote-31) were used to create a more extensive emoji list that resulted in a list containing 751 identifiable emojis. Note that the Unicode consortium has over 2,300 recognized emojis, some of which consist of double character emojis such as the ones shown in Figure 4 with the main emoji, in this case “bearded person” paired with various skin tone colors.

Library_17.1:Users:Guest:Desktop:Screen Shot 2017-04-28 at 1.43.36 PM.pngLibrary_17.1:Users:Guest:Desktop:Screen Shot 2017-04-28 at 1.43.58 PM.pngLibrary_17.1:Users:Guest:Desktop:Screen Shot 2017-04-28 at 1.44.04 PM.pngLibrary_17.1:Users:Guest:Desktop:Screen Shot 2017-04-28 at 1.44.09 PM.pngLibrary_17.1:Users:Guest:Desktop:Screen Shot 2017-04-28 at 1.44.13 PM.png

**Figure 4:** Examples of Double Character Emojis

Lastly, in order to finalize the data processing stage both the text and emoji sentiment scores were calculated. Note that for conducting sentiment analysis each tweet is separated into two distinct data sets – one for text and one for emoji. They are not analyzed as a single entity. Two methods are discussed below as well as the advantages and disadvantages for each method. The first method is the absolute proportional difference method, which looks at how many positive and negative words or emojis exist in a tweet divided by the total number of text or emoji existing in the tweet.[[32]](#footnote-32) This score ranges from 0 to 1. Mathematically this method is calculated by:

The disadvantage of using this method is that a tweet’s score can be heavily affected by non-sentiment-related content. For example if there are words that are not given a sentiment strength in the lexicon used for the analysis they could skew the sentiment strength of the tweet. The same issue can exist for calculating sentiment with emoji since the lexicon only contains 751 emojis out of 2,389. The second method is the relative proportional difference method that calculates a score that ranges from -1 to 1.[[33]](#footnote-33) Mathematically it is calculated by:

Here sentiment is calculated by only using words recognized as having sentiment strength in the lexicons used for this analysis. The disadvantage of using this method is that a sentence’s score may tend to strongly cluster and the end points of the scale given that the tweet content may be primarily or exclusively positive or negative. Despite the second method’s disadvantages, it will be used for this study to calculate sentiment strength and attempts to improve the sentiment strength will be made.

In an attempt to improve the text lexicon used in this study, n-grams were added to account for double negatives such as “not angry” or “wasn’t horrible”. As a first attempt only bi-grams were attempted to see how many bi-grams could be successfully polarized given the Afinn lexicon.

As a second attempt to improve the text lexicon, weighted text frequency was added by looking at an individual word’s inverse document frequency (*idf*), which “decreases the weight for commonly used words and increases the weight for words that are not used very much in a collection of documents.”[[34]](#footnote-34) This addition of *idf* to term frequency (known as *tf-idf*) attempted to adjust for how rarely a word is used within the realm of all tweets used in this stage of analysis – 26,026. As *tf-idf* measures how important a word is to a document in a corpus of documents, this study attempted first to treat each tweet as a single document and all tweets as the document corpus. The result was that the diversity within a single document, or tweet, was so low that every word was given a high *tf-idf* score. In the second attempt, the entire tweet corpus was used and the result was that all of the terms were given *tf-idf* scores that were so small the results were displayed as zero for every word. As a third attempt tweets were categorized into political party reference groups which was used in the analysis below.

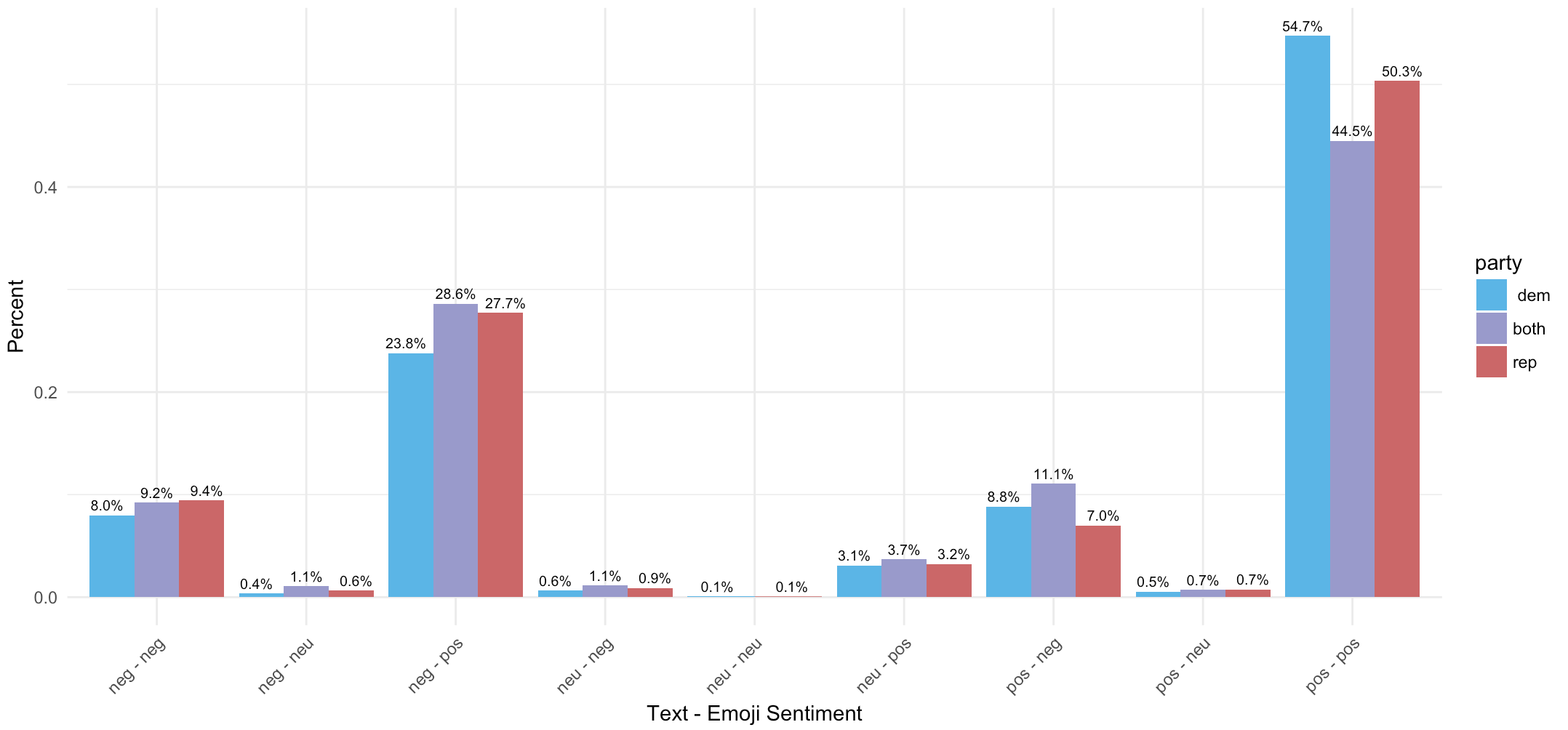
After calculating adjusted weight frequencies for the text corpus attempts were made to integrate this information with the original Afinn lexicon. Multiplying the Afinn sentiment strengths with the new weighted frequencies resulted in the figures depicted below. Since the tweet corpus was broken into three ‘documents’ in order to calculate *tf-idf* – namely Republican, Democrat and both political parties – there was the possibility of having three separate weighted sentiment scores for a single word if it was used in all three political party reference groups. In an effort to normalize any differences, the average of up to three weighted frequencies per word was taken.

# Results

As shown below in Figure 5, 51.1% of tweets contained both positive text and emoji sentiment followed by 26.4% of tweets containing positive emoji and negative text sentiment. The average emoji sentiment was 0.206 while the average for text sentiment was 0.014. When broken down by political party as shown in Figure 6, this trend remains.

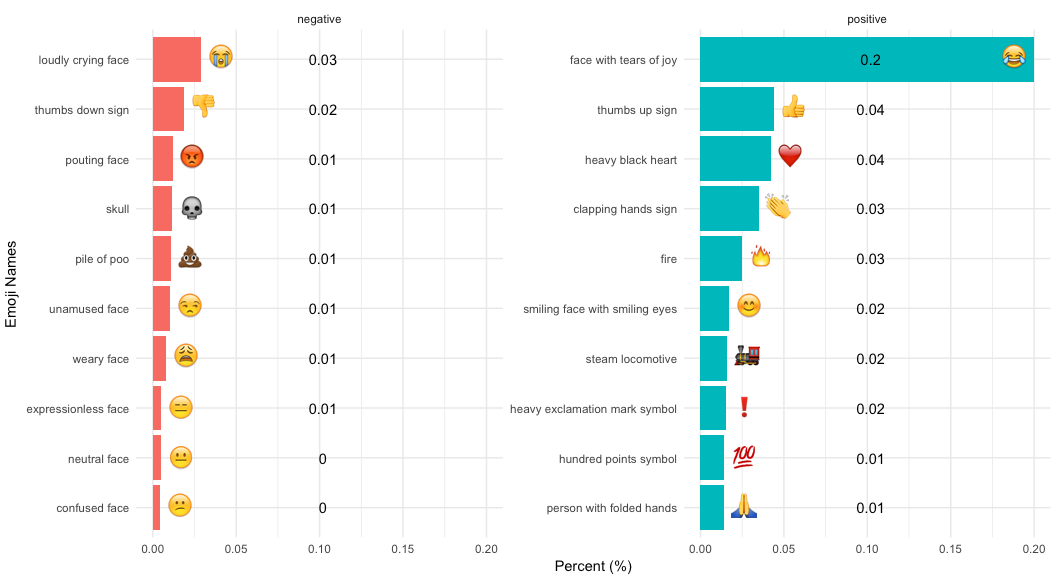
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  | **Text Sentiment** | | | |
|  |  | **Negative** | **Neutral** | **Positive** | **Total** |
| **Emoji Sentiment** | **Negative** | 8.9 | 0.8 | 7.8 | **17.5** |
| **Neutral** | 0.6 | 0.1 | 0.7 | **1.4** |
| **Positive** | 26.4 | 3.2 | 51.6 | **81.2** |
| **Total** | **35.9** | **4.1** | **60.1** | **100** |

**Figure 5:** Text Sentiment by Emoji Sentiment – Percent (%)



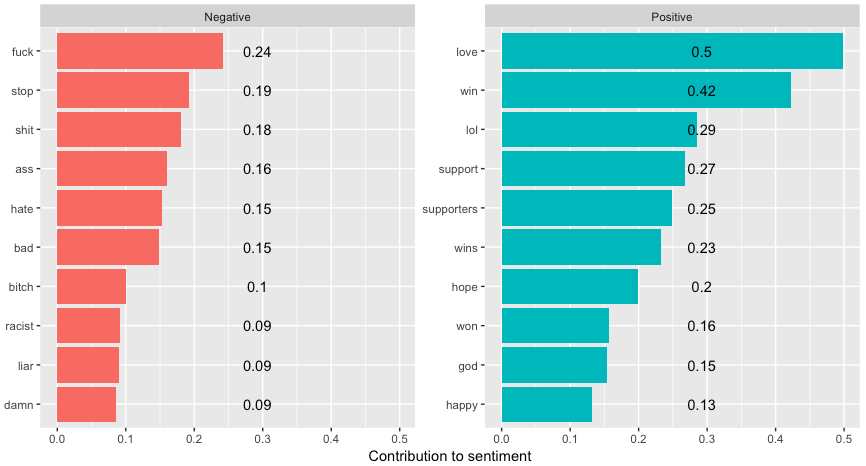
**Figure 6:** Text – Emoji Sentiment by Political Party Reference

When looking at emoji sentiment within this subset, Figure 7 shows that the emoji named“face with tears of joy” was the most commonly used emoji in tweets with some sort of party affiliation (20%). The remaining top five emojis used in tweets areLibrary_17.1:Users:Guest:Desktop:Screen Shot 2017-04-28 at 1.58.47 PM.png “thumbs up” (4.4%), Library_17.1:Users:Guest:Desktop:Screen Shot 2017-04-28 at 1.59.58 PM.png “heavy black heart” (4.2%), Library_17.1:Users:Guest:Desktop:Screen Shot 2017-04-28 at 2.00.25 PM.png “clapping hands” (3.5%) andLibrary_17.1:Users:Guest:Desktop:Screen Shot 2017-04-28 at 2.01.00 PM.png “loudly crying face” (2.9%). Within the top five most used emojis four out of five emojis were polarized as having positive sentiment while the Library_17.1:Users:Guest:Desktop:Screen Shot 2017-04-28 at 2.01.00 PM.pngemoji was polarized as having negative sentiment. Note that there is also a large gap of 15.6 percentage points between the most used emoji and the second most used emoji.



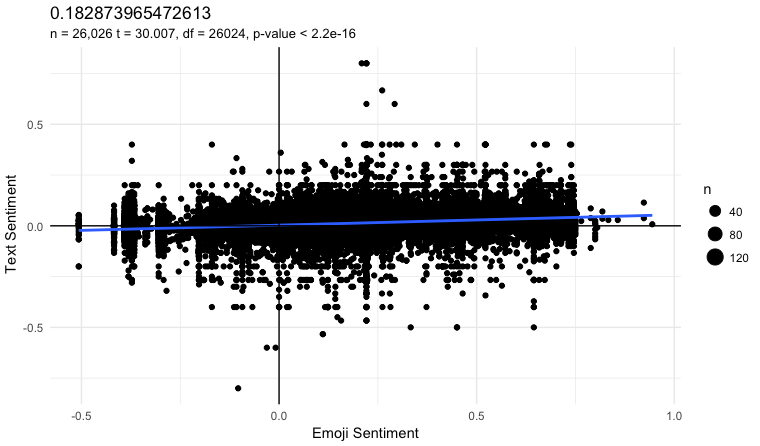
**Figure 7** – Top 20 Emojis used within Tweets with Party References by Sentiment Strength

When looking at the text results we see that the top five most used words were “love” at 50%, followed by “win” at 42%, “lol” at 29%, “support” at 27% and “supporters” at 25% as shown in Figure 8. Since the last two words which have the same stem were not combined, it is one of potentially many examples were stemming is inappropriate or ineffective for natural language processing. Additionally, note the presence of “lol” in the Afinn dictionary with correctly polarized sentiment is potentially a sign of how valuable this lexicon is over other lexicons that do not include such words in their analysis. Had another lexicon been used that doesn’t incorporate common slang or internet words, the third most commonly used word might not have been identified. Looking at the top most used words all five contain positive sentiment. It is important to note that the sixth most common word, “fuck” with 24%, has a negative sentiment strength.



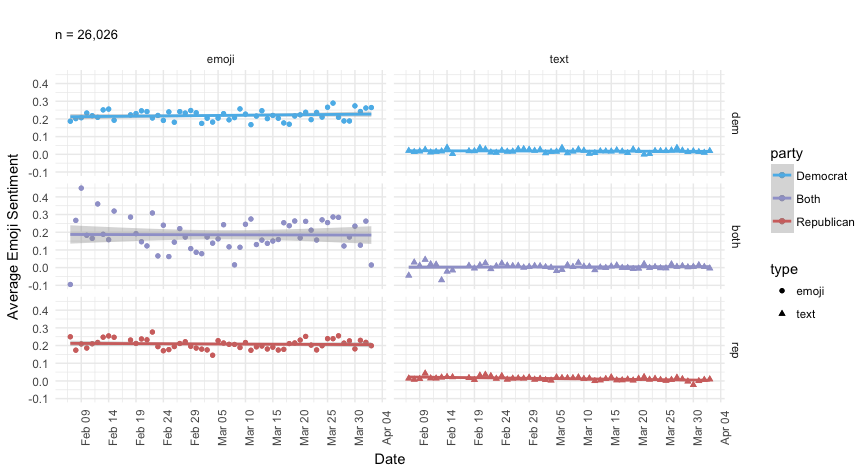
**Figure 8** – Top 20 Words used within Tweets with Party References by Sentiment Strength

Figure 9 looks at the correlation between emoji and text sentiment and shows that there is a correlation of 0.18 between text and emoji sentiment. While this correlation appears quite weak it is important to note that this correlation is statistically significant with a confident interval of 95%. Thus this positive relationship between text and emoji sentiment is not likely to be a result of chance.



**Figure 9** – Sentiment of Text vs. Emoji

Looking at the average sentiment of text and emoji over time by political party, Figure 10 below shows that emoji sentiment remains, on average, positive over time by political party reference. There appears to be a slight increase in positive sentiment over time for tweets referencing democrats compared to the slight increase in negative sentiment over time for tweets referencing republicans. When looking at text sentiment over time the average sentiment remains around zero for all three political party reference groups.



**Figure 10** – Sentiment of Text and Emoji over Time by Political Party

Figure 11 below shows that for emoji sentiment, 81.2% of tweets were categorized as having positive sentiment strength, 17.5% for negative sentiment strength and only 1.3% for neutral sentiment strength. For text sentiment 60% of tweets were categorized as having positive sentiment strength, 35.9% for negative sentiment strength and only 4.1% for neutral sentiment strength.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Positive %** | **Neutral %** | **Negative %** | **Total %** |
| **Emoji** | 81.2 | 1.3 | 17.5 | **100** |
| **Text** | 60 | 4.1 | 35.9 | **100** |

**Figure 11** – Sentiment Strength of All Tweets by Type of Character

Figure 12 shows the sentiment strength of tweets by political party reference and type of characters used. There are more positive emojis in tweets referring to Republicans (49.9%) compared to positive emojis used in tweets referring to Democrats (27.9%). Within the text sentiment we see that there are more positive tweets referring to Republicans (35.7%) negative sentiment for Republicans (23.2%). For additional information about the distribution of tweets refer to Appendix E Figures 3.a and 3.b which show a right skew to the distribution of emoji sentiment and what appears to be a normal distribution for text sentiment.

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | **Emoji** | **Text** |
| **Party** | **Sentiment Strength** | **Percent** | **Percent** |
| **Democrat** | **+** | 27.9 | 21.9 |
|  | **0** | 0.3 | 1.3 |
|  | **-** | 6 | 11 |
| **Both** | **+** | 3.4 | 2.5 |
|  | **0** | 0.1 | 0.2 |
|  | **-** | 0.9 | 1.7 |
| **Republican** | **+** | 49.9 | 35.7 |
|  | **0** | 0.9 | 2.6 |
|  | **-** | 10.6 | 23.2 |
| **Total** |  | **100** | **100** |

**Figure 12** – Sentiment Strength of Tweets by Political Party and Type of Character

In the first attempt to improve the Afinn text lexicon, bi-grams were looked at. Classifying the text corpus into bi-grams resulted in 215,087 unique bi-grams. Figure 13 shows that the most common bi-gram is “donald trump” at 24.84% followed by “bernie sanders” at 0.84%. Note that the top ten bi-grams are either references to the individual people running in the presidential primaries or they contain either one or two common stop words typically removed when dealing with individual words.

|  |  |
| --- | --- |
| Word | Percent |
| donald trump | 24.84 |
| bernie sanders | 0.84 |
| ted cruz | 0.56 |
| vote for | 0.5 |
| is a | 0.44 |
| trump is | 0.44 |
| in the | 0.44 |
| hillary clinton | 0.42 |
| of the | 0.33 |
| is the | 0.33 |

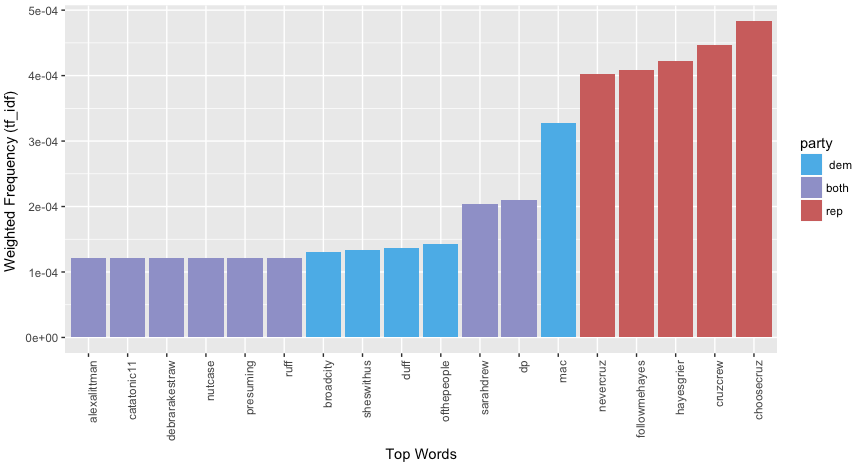
**Figure 13** – Top Ten Most Common Bi-Grams

When the 215,087 bi-grams were polarized only nine bi-grams were successfully computed as seen in Figure 14. Given that only 0.0042% of bi-grams were polarized it is not feasible to include polarized bi-grams into this model.

|  |  |  |
| --- | --- | --- |
| Word | Score | Count |
| can’t stand | -0.6 | 32 |
| fed up | -0.6 | 26 |
| not good | -0.4 | 16 |
| not working | -0.6 | 6 |
| screwed up | -0.6 | 2 |
| some kind | 0.0 | 2 |
| dont like | -0.4 | 2 |
| no fun | -0.6 | 1 |
| right direction | 0.6 | 1 |

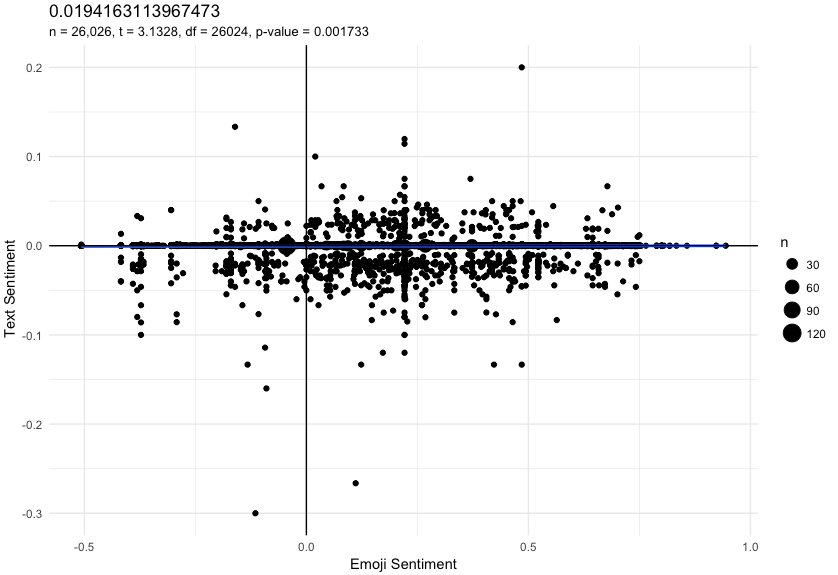
Figure 14 – Polarized Bi-Grams

As a second attempt to improve the text lexicon, tweets were categorized into political party reference groups, which provided more reasonable *tf-idf* results as shown in Figure 15. The top five most frequently weighted words are all from Republican references – “choosecruz”, “cruzcrew”, “hayesgrier”, “followmheyes” and “nevercruz”.



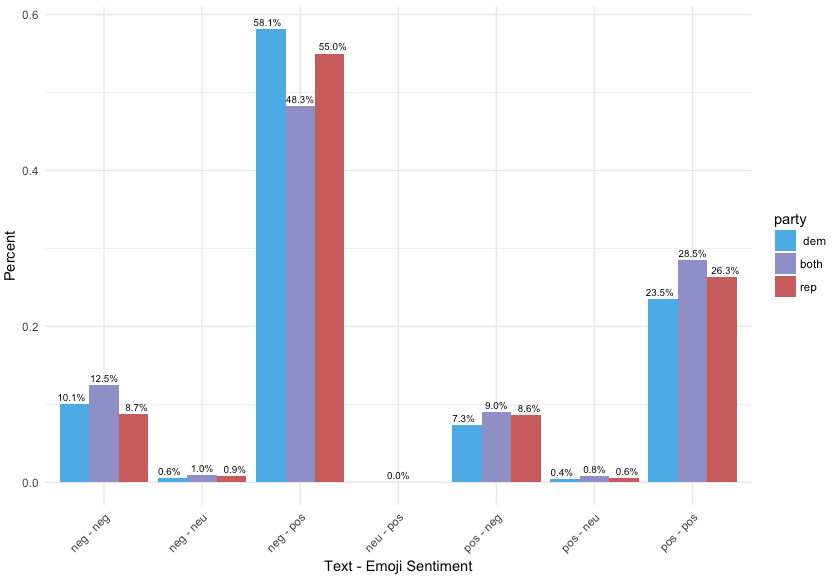
**Figure 15** – Top Weighted Frequency of Words by Political Party

Figure 16 shows an updated version of the relationship between text and emoji sentiment. Compared to Figure 9 above, the correlation is weaker at 0.019 than the earlier correlation of 0.18. This new value is still statistically significant but its p-value is much larger (0.0017 compared to 2.2e-16).



**Figure 16**–Weighted Sentiment of Text vs. Emoji

After adding this weighted frequency, the most common tweet pairing for text and emoji sentiment moved from being positive text – positive emoji in Figure 9 to negative text – positive emoji as shown in Figure 17. Given that the range *tf-idf* for identified words was -0.0013340622 to 0.0001218875 with a median of -1.91046e-05, it appears that more words were given negative weight than positive weight.



**Figure 17** –Weighted Text- Emoji Sentiment by Political Party Reference

|  |  |
| --- | --- |
| Top Emoji Tweets | |
| Top Positive Tweets | **Score** |
| Macintosh HD:Users:alexandraplassaras:Desktop:Screen Shot 2017-04-20 at 11.46.09 PM.png | 0.94 |
| Macintosh HD:Users:alexandraplassaras:Desktop:Screen Shot 2017-04-20 at 11.46.46 PM.png | 0.92 |
| Macintosh HD:Users:alexandraplassaras:Desktop:Screen Shot 2017-04-20 at 11.47.22 PM.png | 0.92 |
| Macintosh HD:Users:alexandraplassaras:Desktop:Screen Shot 2017-04-20 at 11.47.43 PM.png | 0.86 |
| Macintosh HD:Users:alexandraplassaras:Desktop:Screen Shot 2017-04-20 at 11.48.16 PM.png | 0.83 |
| Top Negative Tweets | **Score** |
| Macintosh HD:Users:alexandraplassaras:Desktop:Screen Shot 2017-04-20 at 11.50.29 PM.png | -0.51 |
| Macintosh HD:Users:alexandraplassaras:Desktop:Screen Shot 2017-04-20 at 11.50.46 PM.png | -0.51 |
| Macintosh HD:Users:alexandraplassaras:Desktop:Screen Shot 2017-04-20 at 11.51.05 PM.png | -0.51 |
| Macintosh HD:Users:alexandraplassaras:Desktop:Screen Shot 2017-04-20 at 11.51.28 PM.png | -0.51 |
| Macintosh HD:Users:alexandraplassaras:Desktop:Screen Shot 2017-04-20 at 11.51.45 PM.png | -0.51 |

**Figure 18** – Top Emoji Sentiment Tweets

Figures 18 and 19 showcase the tweets with the top positive and negative sentiment scores for both emojis and text. From Figure 18 we can see that the most positive sentiment score for tweets with emojis was 0.94 for a tweet that contained  which is “Japanese Dolls”. Another emoji associated with the top positive tweets was  (“bell”). For the top negative tweets, one emoji stands out as having the most negative score -  known as the “imp”. Note however that for negative sentiment scores for emojis that the score remains the same for tweets that contain multiples of the same emoji. Thus, a tweet that contains  is given the same score as a tweet that contains only a single . This is also something seen in positive sentiment scores for emojis.

|  |  |  |  |
| --- | --- | --- | --- |
| Afinn Lexicon |  | Weighted Afinn Lexicon |  |
| Top Positive Tweets | **Score** | **Top Positive Tweets** | **Score** |
| Macintosh HD:Users:alexandraplassaras:Desktop:Screen Shot 2017-04-21 at 12.02.04 AM.png | 0.80 | Macintosh HD:Users:alexandraplassaras:Desktop:Screen Shot 2017-04-24 at 1.31.22 PM.png | 0.20 |
| Macintosh HD:Users:alexandraplassaras:Desktop:Screen Shot 2017-04-21 at 12.01.51 AM.png | 0.80 | Macintosh HD:Users:alexandraplassaras:Desktop:Screen Shot 2017-04-24 at 1.31.10 PM.png | 0.13 |
| Macintosh HD:Users:alexandraplassaras:Desktop:Screen Shot 2017-04-21 at 12.01.36 AM.png | 0.80 | Macintosh HD:Users:alexandraplassaras:Desktop:Screen Shot 2017-04-24 at 1.30.54 PM.png | 0.12 |
| Macintosh HD:Users:alexandraplassaras:Desktop:Screen Shot 2017-04-21 at 12.01.23 AM.png | 0.67 | Macintosh HD:Users:alexandraplassaras:Desktop:Screen Shot 2017-04-24 at 1.30.37 PM.png | 0.11 |
| Macintosh HD:Users:alexandraplassaras:Desktop:Screen Shot 2017-04-21 at 12.01.10 AM.png | 0.60 | Macintosh HD:Users:alexandraplassaras:Desktop:Screen Shot 2017-04-24 at 1.25.09 PM.png | 0.10 |
| Top Negative Tweets | **Score** | **Top Negative Tweets** | **Score** |
| Macintosh HD:Users:alexandraplassaras:Desktop:Screen Shot 2017-04-21 at 12.06.01 AM.png | -0.53 | Macintosh HD:Users:alexandraplassaras:Desktop:Screen Shot 2017-04-24 at 1.31.35 PM.png | -0.13 |
| Macintosh HD:Users:alexandraplassaras:Desktop:Screen Shot 2017-04-21 at 12.06.16 AM.png | -0.53 | Macintosh HD:Users:alexandraplassaras:Desktop:Screen Shot 2017-04-24 at 1.31.47 PM.png | -0.13 |
| Macintosh HD:Users:alexandraplassaras:Desktop:Screen Shot 2017-04-21 at 12.06.30 AM.png | -0.60 | Macintosh HD:Users:alexandraplassaras:Desktop:Screen Shot 2017-04-24 at 1.32.01 PM.png | -0.16 |
| Macintosh HD:Users:alexandraplassaras:Desktop:Screen Shot 2017-04-21 at 12.06.43 AM.png | -0.60 | Macintosh HD:Users:alexandraplassaras:Desktop:Screen Shot 2017-04-24 at 1.32.15 PM.png | -0.27 |
| Macintosh HD:Users:alexandraplassaras:Desktop:Screen Shot 2017-04-21 at 12.06.58 AM.png | -0.80 | Macintosh HD:Users:alexandraplassaras:Desktop:Screen Shot 2017-04-24 at 1.32.27 PM.png | -0.30 |

**Figure 19** – Top Text Sentiment Tweets

Figure 19 shows that top positive and negative tweets for text sentiment with the original Afinn lexicon examples shown on the left and the new weighted Afinn lexicon examples shown on the right. Within the original lexicon examples we see that words like “win”, “LMAO”, “awesome”, “congratulations” and “love” are all associated with highly positive scores while curse words like “fucking”, “dickhead”, “bitch”, “hell” and “bullshit” are all associated with very negative scores. After applying a weighted frequency score to the tweets we see that less common but still positive words like “cleaver”, “hugs”, “great” and “ecstatic” are found in highly positive tweets. Of the top negatively scored tweets we see that “dickhead” remains highly negative and uncommon enough to be given a large weight as well as the words “jackass” and “heartbroken”. It is useful to note however that the lexicon used here is not 100% accurate as the fifth most positive tweet with the weighted score gave the word “ethical” a highly positive score whereas the actual sentiment of this tweet appears less than positive and more skeptical of Bernie Sanders. This is likely due to the fact that all punctuation was removed from the text corpus that was used to calculate sentiment.

# Discussion

## Results

As discussed previously the three hypotheses for this study were 1) that there would be a significant difference between sentiment analysis of text compared to the sentiment analysis of emojis, 2) tweets referencing Republican candidates would contain more negative sentiment for emojis than tweets referencing Democratic candidates and 3) tweets that mention Democratic candidates would be more likely to have higher negative text sentiment than tweets mentioning Republican candidates. In regards to the first hypothesis there is a weak positive correlation between text and emoji sentiment, which goes against this hypothesis. Before any analysis was conducted, this paper believed that comparing emoji sentiment with text sentiment could be one way for researchers to identify sarcasm so long as the first hypothesis was met. Instead, this paper finds that on average users who tweet are more likely to have emojis and text with similar sentiment strength. Perhaps with a more robust text lexicon, results will be found that prove the first hypothesis but until then it appears looking at the differences between text and emoji usage might not be the best measure to identify sarcasm with. As discussed below, an improved text analysis will most likely provide better insight as this paper’s model does not account for double negatives, bi-grams or spelling errors – all valuable attributions for analyzing text.

As Figure 12 showed above, in general tweets referring to Republicans appear to have more negative sentiment for emojis at 10.6% than tweets referring to Democrats at 6%, which is what the second hypothesis of this paper stated. The same figure also shows that tweets referring to Republicans appear to have more negative sentiment for text at 23.3% compared to tweets referring to Democrats at 11%. This negates what the third hypothesis stated.

Throughout the presidential election, the Republican party’s investment on social media may have had a much larger impact on the outcome of the election. A Bloomberg article that broke down spending by political party in the presidential election stated that Trump’s digital strategy may have been decisive in his victory. Per the article, a Trump campaign official stated that social media platforms were used to target “specific groups of Clinton backers with negative ads on social media to lower Democratic turnout”[[35]](#footnote-35). The article continues by highlighting the Clinton campaign’s focus on television ads and get-out-the-vote campaigns but does not mention much focus on social media platforms. Thus, the results found in this paper may not be generalizable to other large tweet corpora that are not political in nature.

Another takeaway from this study is that users tended to choose words with extreme sentiment. The most commonly used word in the entire corpus was “love” at 50% and the most common negative word was “fuck” at 24%. One could argue that these words are associated with strong emotions. Again while this might not be indicative of all tweets, it is useful to note that Twitter users – or potentially users on other social media platforms – might have higher tendencies to communicate in extremes or by using words with high emotional sentiment to them. In regards to emojis used, the clear winner of most common emoji used is the face with tears of joy at 20%. This might be because that emoji has an ambiguous meaning. It could symbol laughing so much that tears stream down one’s face or it could be a sarcastic emoji expressing laughing but crying at the same time.

## Limitations

Limitations to this study include the potential for sampling error, difference between British and American English, time restraints, misunderstanding of emoji, lack of context for emoji analysis, accuracy of sentiment analysis, the use of non-standard words used during this campaign cycle, the *idf* used to modify the Afinn lexicon and the accuracy of the text lexicon. The first limitation to this study is the large possibility of sampling error. According to Pew Research Center in 2015 it was estimated that only 23% of all internet users and 20% of the entire adult population in the US use Twitter.[[36]](#footnote-36) As such this data set is neither representative of all Americans nor is it representative of Democrats or Republicans. Another study conducted by Pew Research Center estimated in 2012 that of the 16% of internet users that used Twitter, 12% of users were estimated to be Republicans and around 18% were estimated to be Democrats.[[37]](#footnote-37) Additionally, this study only looks at Tweets and not other types of social media or e-communication like Facebook, text messages or emails etc. Since this data set is not representative of the population intended to be analyzed – Americans who use electronic communication – there is potential for selection bias.

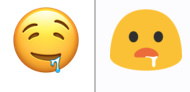
The second limitation of this study is that the lexicon used to analyze emoji sentiment was created for a separate study that looked at the sentiment of emojis in the UK. Jack *et al.* suggests that people’s interpretation of facial expressions and thus emojis differ between cultures.[[38]](#footnote-38) Another study that looked at emoji usage of smartphone users ranked the top 10 emojis in the top 10 countries to see which emojis were most often used in different places.[[39]](#footnote-39) France was the only country in which all 10 emojis that were most used contained a heart somewhere in the emoji (See Appendix E for the table results). Park *et al.* discusses the fact that “easterners and westerners prefer different style of emoticons”[[40]](#footnote-40) but perhaps there are more differences between countries than just being considered eastern or western. As a result, the true sentiment of the emojis in this American data set may not be fully represented when using a lexicon built for British emoji use. However the same study that showed the top 10 French heart emojis result also made the claim that “countries sharing similar emoji usage patterns are more likely to share common language or geo-region.”[[41]](#footnote-41) Perhaps, then, the difference between using a British based lexicon for an American data set might not have that much of an effect but it is something to consider for this study.

The third limitation is one of time constraints. Given that this data set contains only 56 days worth of data and the fact that this data was specifically capturing only tweets pertaining to the 2016 primary elections, Lu *et al.* suggests in their paper that these events may have led to “unrepresentative user moods and behaviors” that could have affected how users chose emojis.[[42]](#footnote-42)

The fourth limitation of this study is that it does not account for the misunderstanding of emojis by various users. Tigwell and Flatla state that two common reasons for users to misunderstand emoji are 1) the definition and use of emoji and 2) the different emoji designs on different platforms.[[43]](#footnote-43) People’s opinion on how emojis should be used and what they represent can vary greatly. For example, what exactly is the emoji in Figure 20 below and should it be used as a positive emoji or a negative emoji? Additionally, depending on whether a person is using an iOS or an Android phone, emojis can look very different and perhaps even have different meaning entirely. Figure 21below shows the same Unicode emoji that looks slightly different between iOS and Android platforms. It is possible that a user might think the Android emoji on the right is more negative than the iOS emoji is more positive. Both emoji have the label “drooling face” but the emotions they convey might not be the same for every user.



**Figure 20** – “Smiling face with open mouth & cold sweat” Emoji



**Figure 21** – “Drooling face” Emoji (iOS - left, Android - right)

The fifth limitation of this study is that context is not accounted for in the sentiment analysis for the emoji. The analysis of both the text and the emoji are done in two different silos and as a result the sentiment strength of the overall tweet is not accounted for. It could be that a tweet that has very negative text but very positive emojis may actually be either negative or positive as a whole but since the text and emoji are not calculated together we would not be able to discern this from the analysis conducted in this study alone.

The sixth limitation is the failure of sentiment analysis to account for sarcasm. Sarcasm in general is difficult to analyze, for both humans and machines. In order to identify and understand sarcasm the context of the situation, cultural norms and topical information must be known.[[44]](#footnote-44) This amount of information is almost impossible for a machine to account for and then analyze. While algorithms have been created to detect high success rates of sarcasm as in the French company Spotter[[45]](#footnote-45), the analysis used in this paper is not as robust in its analysis. Thus the sentiment strength of a tweet’s text might not be accurately depicting a user’s sentiment.

The seventh limitation that might differ between data sets is that there may exist in the text corpus words that are not recognized in the text lexicon used that may be of importance to this analysis such as balloonomania, nanity, questmonger[[46]](#footnote-46), words not found in the English dictionary such as braggadocious[[47]](#footnote-47) or internet slang such as gr8, jk or nsfw[[48]](#footnote-48) which were all words used during the primaries either within the data for this study or from candidates themselves. The Afinn lexicon does include some slang words as well as curse words but this lexicon was created in 2011 so more recent internet slang may not be accounted for.

The eighth limitation is the term frequency – inverse document frequency (*tf-idf*)that was used to modify the original sentiment scores calculated from the Afinn lexicon. In order to calculate the *idf* the entire text corpus has to be separated into what are known as documents. If they are not subset into these documents then the *idf* term for every word becomes so small it is indistinguishable from zero. In this study, the documents were the political party references of Democrat, Republican or both parties referenced in the same tweet. This subsetting did not use any classification algorithms and was created using a small list of words known to be associated with each political party – such as the name of the party or shorthand for said party and the candidates running for each party. Thus this separation might not be entirely accurate and could also have prevented many tweets from being placed in either three categories. If for example there were tweets referencing donkeys and elephants – the symbols for the political parties – but no reference to the party or a candidate specifically, then that tweet would have been categorized in the other group and not analyzed at all since no discernable party was detected.

The last limitation discussed in this paper is the accuracy of the current text lexicon used in this study. The Afinn lexicon used in this study doesn’t account for spellchecking errors, the semantics or subtle shades of meaning behind words that seemingly mean the same thing or

booster words such as “very”, “really” or “extremely”. Additionally the Afinn lexicon has a bias towards negative words 65% compared to positive words 35%.[[49]](#footnote-49) This fact could affect the calculation of tweet sentiment and is something to consider when viewing the results of this paper.

# Conclusion

While the context about what the data is and where it came from is important to consider when reading the findings of this paper, the main purpose of this paper was attempting to understand the relationship between text and emojis. From the analysis presented in this paper we see at least a weak positive relationship that is statistically significant between text and emoji usage in Twitter data. This information might be useful for future political campaign analysis to consider when evaluating a candidates success in a race or when attempting to gain voter sentiment during a campaign as analysts are able to capture tweets in real time. More work, however should be done to improve on the work done in this paper.

Creating a classification algorithm for political party reference would be a useful next step as it is highly unlikely that 1) Twitter will ask users to self report which political party they affiliated with, 2) that enough users would voluntarily associate this information with their Twitter accounts and 3) that this information could be verified as being accurate. In regards to the lexicons used in this study, more robust lexicons for text should in theory improve the sentiment calculation. As mentioned above the Afinn lexicon does not account for double-negatives, sarcasm, spell-check errors, booster words or semantics. Within the emoji lexicon, tweets that contain varying levels of the same emoji are all associated with one score. For example, a tweet that contains  is given the same score as a tweet that contains only a single .

In regards to the calculation of sentiment, an alternative method should be looked at in further detail. Specifically a logit scale, which ranges from positive to negative infinity, might be a better way of calculating sentiment strength. Mathematically this can be calculated by:

Where the 0.5 helps smooth the results and prevents log(0) (or an undefined result) from occurring. The benefit of this method is that it is symmetric around zero and when compared to the other two methods has the smoothest properties.[[50]](#footnote-50) Lastly, beyond the scope of this data set in the future it would be useful to expand this research by looking at other large data sets of social media or mobile communication data to see whether the sentiment strength patterns discussed in this paper follow the similar patterns.

# Appendices

### Appendix A – QMSS G4063 Information

Below are links to Data Processing and Data Visualization’s github page as well as a link to the original JSON data containing all tweets scraped.

* <https://github.com/hassanpour/QMSS_G4063>
* [https://www.dropbox.com/sh/zyy9tsvibrl4d63/AAAQ6D3h0Kksxb8EeVH2RSSAa/tweets\_geo\_all.json?dl=0#](https://www.dropbox.com/sh/zyy9tsvibrl4d63/AAAQ6D3h0Kksxb8EeVH2RSSAa/tweets_geo_all.json?dl=0)

### Appendix B – All variables associated with a single tweet

> colnames(tweetsUS)

[1] "X.1" "id\_str"

[3] "idx" "text"

[5] "created\_at" "screen\_name"

[7] "user\_lang" "truncated"

[9] "retweeted" "favorite\_count"

[11] "verified" "user\_id\_str"

[13] "source" "followers\_count"

[15] "in\_reply\_to\_screen\_name" "location"

[17] "retweet\_count" "favorited"

[19] "utc\_offset" "statuses\_count"

[21] "description" "friends\_count"

[23] "user\_url" "geo\_enabled"

[25] "in\_reply\_to\_user\_id\_str" "lang"

[27] "user\_created\_at" "favourites\_count"

[29] "name" "time\_zone"

[31] "in\_reply\_to\_status\_id\_str" "protected"

[33] "listed\_count" "place\_lon"

[35] "expanded\_url" "place\_id"

[37] "full\_name" "lat"

[39] "country\_code" "place\_name"

[41] "url" "country"

[43] "lon" "place\_type"

[45] "place\_lat" "X"

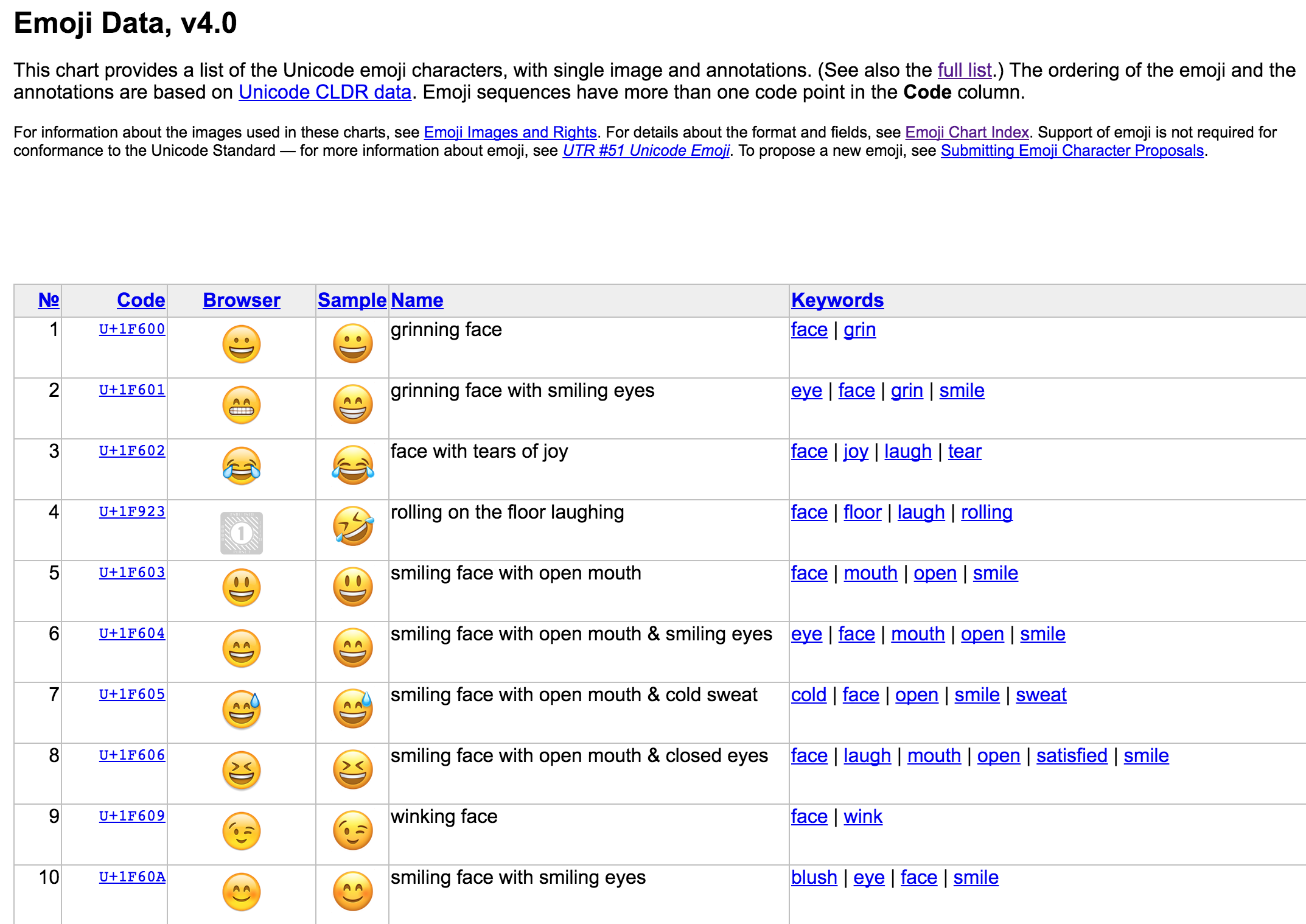
[47] "Y" "STATEFP"

[49] "NAME" "COUNT"

### Appendix C – Unicode List of Emoji

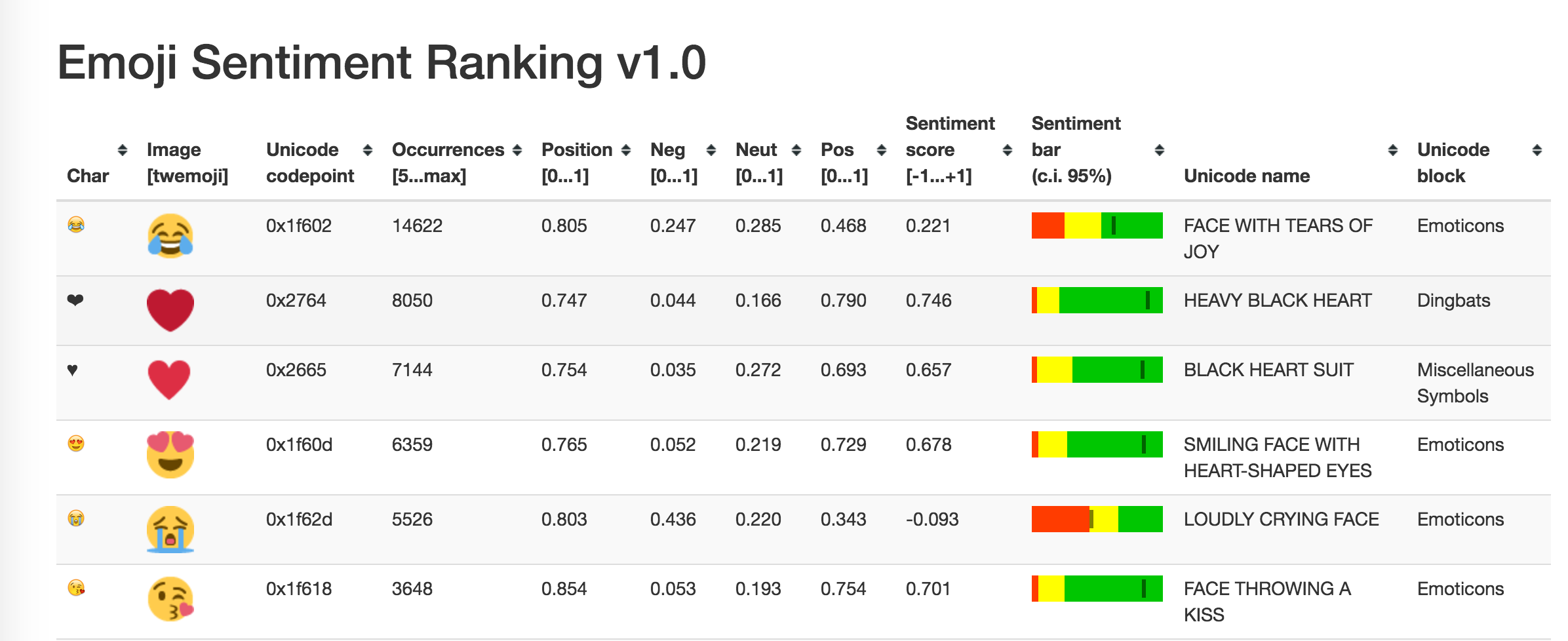
Below is a screenshot of the Unicode Consortium’s list of all available emojis. To view more, visit:

http://unicode.org/emoji/charts/emoji-list.html

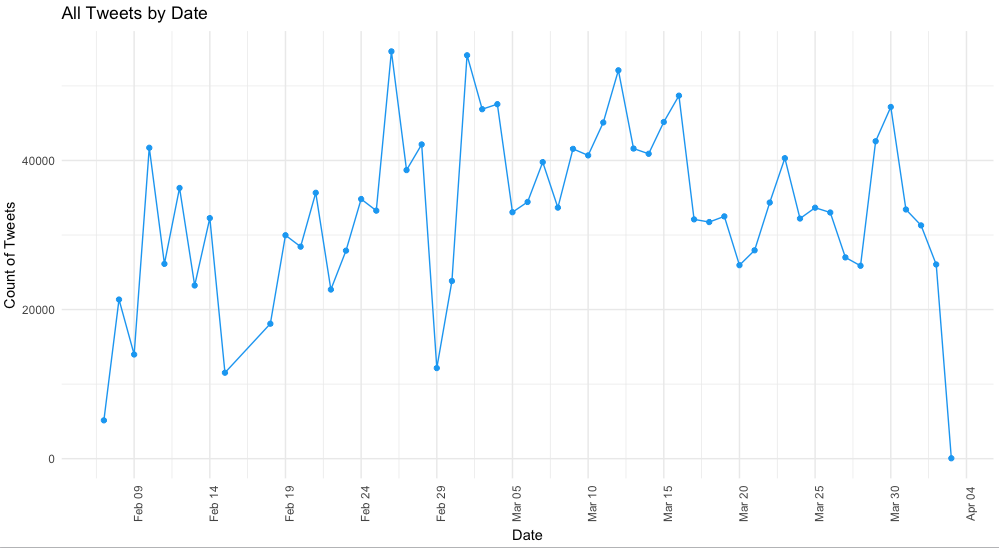


### Appendix D – Emoji Sentiment Lexicon

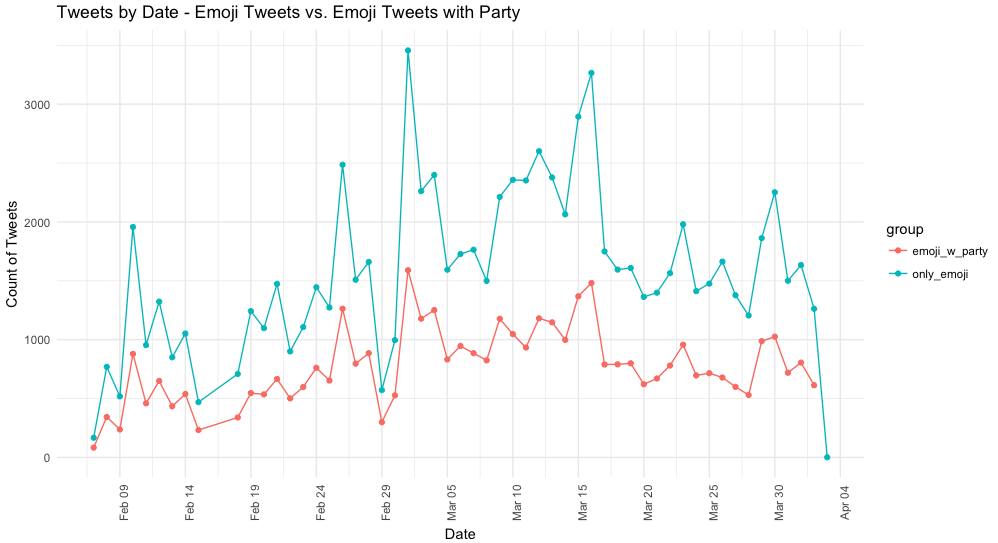
Below is a screenshot of the information found on <http://kt.ijs.si/data/Emoji_sentiment_ranking/>.



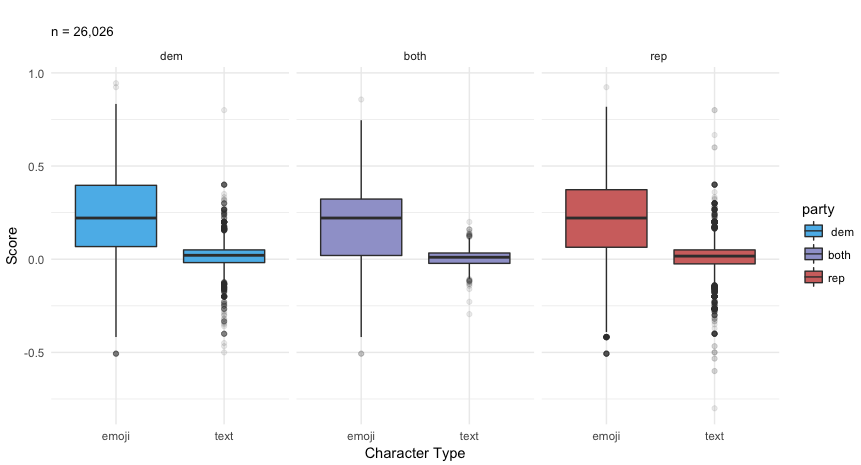
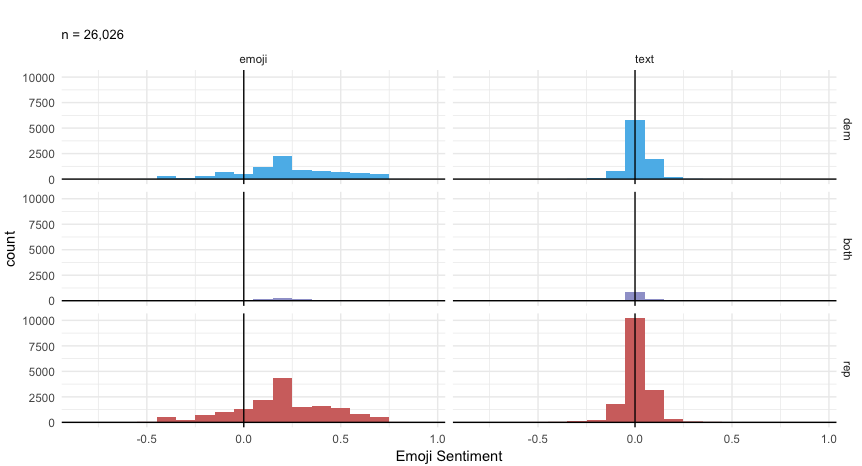
### Appendix E – Additional Visualizations



**Figure 1.a** – Count of Tweets by Day



**Figure 1.b** – Count of Tweets by Day – Emoji Tweets vs. Emoji Tweets with Political Party References



**Figures 3.a & 3.b** – Sentiment of Text and Emoji over Time by Political Party (Left – Histogram, Right - Boxplot)



**Figure 4** – Most common words found in corpus (N > 200)

# References

--. "Author Translates All of 'Alice in Wonderland' into Emojis." *Creators*. N.p., n.d. Web. 02 Dec. 2016.

--. “Emoji Dick.” n.d. Web. 16, Oct. 2016.

--. “Emoji Report.” *Emogi Research Team.* 2015. 21, Nov. 2016.

Allison, Bill, Mira Rojanasakul, Brittany Harris, and Cedric Sam. "Tracking the 2016 Presidential Money Race." *Bloomberg.com.* Bloomberg, 09 Dec. 2016. Web. 28 Apr. 2017.

Boia, M., Faltings, B., Musat, C. C., and Pu, P. 2013. A : ) is worth a thousand words: now people attach sentiment to emoticons and words in Tweets. In Proceedings of the 7th International Conference on Weblogs and Social Media, ICWSM2013. 345-350.

Brown, L. “Twitter - 30 Must-Know Twitter Abbreviations and Acronyms.” 2014. Web. 25, Nov. 2016.

Doble, A. "UK's Fastest Growing Language Is... Emoji - BBC Newsbeat." *BBC News*. BBC, 19 May 2015. Web. 28 Apr. 2017.

Duggan, M. “The Demographics of Social Media Users.” 19, Aug. 2015. Web. 02, Dec. 2016.

Foran, C. “The 'Never Clinton' Campaign. 5”, May 2016. Web. 01, Dec. 2016.

Grady, S. “How can emoji analysis improve results in social analytics and social listening tools?” n.p. 13, May 2016. Web. 12, Nov. 2016.

Jack, R. E., Blais, C., Scheepers, C., Schyns, P. G., and Caldara, R. “Cultural confusions show that facial expressions are not universal.” *Current Biology* 19, 18. 2009. 1543-1548.

Kleinman, Z. “Authorities 'use analytics tool that recognises sarcasm'” 2013. Web. 02, Dec. 2016.

Knapp, M., and Hall, J. A. Nonverbal communication in human interaction (7th ed.). Cengage Learning. 2010.

Laurenancona. "Laurenancona/twimoji." *GitHub*. N.p., 11 July 2015. Web. 28 Apr. 2017.

Liu, Bing. "Sentiment Analysis and Subjectivity." An chapter in Handbook of Natural Language Processing, Second Edition, (editors: N. Indurkhya and F. J. Damerau), 2010.

Lowe, William. Benoit, Kenneth. Mikhaylov, Slava. and Laver, Michael. "Scaling Policy Preferences From Coded Political Texts." *Legislative Studies Quarterly* 26(1, 2) 2011. 123-155.

Lu, X., Ai, W., Liu, X., Li, Q., Wang, N., Huang, G., and Mei, Q. 2016. Learning from the ubiquitous language: an empirical analysis of emoji usage of smartphone users. In Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '16). ACM, New York, NY, USA, 770-780.

Maynard,, D., & Greenwood, M. A. (n.d.). “Who cares about sarcastic tweets? Investigating the impact of sarcasm on sentiment analysis” n.d. Web. 01, Dec. 2016.

Miller, H., Thebault-Spieker, J., Chang, S., Terveen, L. and Hecht, B. ‘"Blissfully Happy" or "Ready to Fight": Varying Interpretations in Emoji. the 10th Conference on Web and Social Media’ n.p. 2016.

Nielsen, Finn Årup. "AFINN." *AFINN*. Informatics and Mathematical Modelling, Technical University of Denmark, Mar. 2011. Web. 28 Jan. 2017.

Nielsen, Finn Årup. "AFINN: A New Word List for Sentiment Analysis on Twitter." *Finn Årup Nielsen's Blog.* N.p., 25 Aug. 2015. Web. 28 Mar. 2017.

Nielsen, Finn Årup. "A New ANEW: Evaluation of a Word List for Sentiment Analysis in Microblogs." (n.d.): n. pag. Dtu Informatics, Technical University Of Denmark, Lyngby, Denmark. Web. 28 Mar. 2017.

Novak, K., Smailović, P., Sluban, J., Mozetič, B. Sentiment of Emojis. PLoS ONE 10(12). 2015.

Pang, B., Lee, L., and Vaithyanathan. S. “Thumbs up? Sentiment classification using machine learning techniques.” In Proceedings of the Conference on Empirical Methods in NLP, 2002. Philadelphia, PA. 79–86.

Park, J., Barash, V., Fink, C., and Cha, M.. 2013. “Emoticon style: interpreting differences in emoticons across cultures.” *In Proceedings of the Seventh International Conference on Weblogs and Social Media,* ICWSM 2013. 2013.

Prismoji. "PRISMOJI/emojis." *GitHub*. N.p., n.d. Web. 28 Apr. 2017.

Rothenberg, Matthew. "Emojitracker: Realtime Emoji Use on Twitter." *Emojitracker: Realtime Emoji Use on Twitter*. N.p., n.d. Web. 28 Apr. 2017.

Schott, B. “Ben Schott on the Strange New Vocabulary of the 2016 Election.” 2016. Web. 02, Dec. 2016.

Shafer, J. “Donald Trump Talks Like a Third-Grader. Retrieved” 13, Aug. 2015. Web. 01, Dec. 2016.

Silge, Julia, and David Robinson. *Text Mining with R*. N.p., 23 Apr. 2017. Web. 28 Apr. 2017.

Smith, A. “Civic Engagement in the Digital Age.” 25, Apr. 2013. Web. 30, Nov. 2016.

Stack, L. “‘Braggadocious?’ In Trump and Clinton Debate, Words That Sent Viewers to the Dictionary.” 2016. Web. 02, Dec. 2016.

Taboada, M., Brooke, J., Tofiloski, M., Voll, K., & Stede, M. Lexicon-Based Methods for Sentiment Analysis. Computational Linguistics, 37(2), 2011. 267-307.

Tigwell, G. W., and Flatla, D. R..“Oh that's what you meant!: reducing emoji misunderstanding.” In Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct (MobileHCI '16). 2016. ACM, New York, NY, USA, 859-866.

Tufts, Chris. "Twimoji: Identifying Emoji in Tweets." *Mining the Details.* N.p., 24 July 2015. Web. 28 Apr. 2017.

Turney, Peter. “Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews.” In Proceedings of 40th Meeting of the Association for Computational Linguistics, 2002. Philadelphia, PA. 417–424.

Walther, J. and D’Addario, K. P..The impacts of emoticons on message interpretation in computer-mediated communication. Social Science Computer Review 5. 2. 2003. 119-134.

Yan, H., Sgueglia, K., & Walker, K. 'Make America White Again': Hate speech and crimes post-election. 29. Nov. 2016. Web. 01, Dec. 2016.

1. "Author Translates All of 'Alice in Wonderland' into Emojis." *Creators*. N.p., n.d. Web. 02 Dec. 2016. [↑](#footnote-ref-1)
2. “Emoji Dick.” n.d. Web. 16, Oct. 2016. [↑](#footnote-ref-2)
3. “Emoji Report.” *Emogi Research Team.* 2015. 21, Nov. 2016. [↑](#footnote-ref-3)
4. Doble, A. "UK's Fastest Growing Language Is... Emoji - BBC Newsbeat." *BBC News*. BBC, 19 May 2015. Web. 28 Apr. 2017. [↑](#footnote-ref-4)
5. Knapp, M., and Hall, J. A. Nonverbal communication in human interaction (7th ed.). Cengage Learning. 2010. [↑](#footnote-ref-5)
6. Grady, S. “How can emoji analysis improve results in social analytics and social listening tools?” n.p. 13, May 2016. Web. 12, Nov. 2016. [↑](#footnote-ref-6)
7. Miller, H., Thebault-Spieker, J., Chang, S., Terveen, L. and Hecht, B. ‘"Blissfully Happy" or "Ready to Fight": Varying Interpretations in Emoji. the 10th Conference on Web and Social Media’ n.p. 2016. [↑](#footnote-ref-7)
8. Walther, J. and D’Addario, K. P..The impacts of emoticons on message interpretation in computer-mediated communication. Social Science Computer Review 5. 2. 2003. 119-134. [↑](#footnote-ref-8)
9. Novak, K., Smailović, P., Sluban, J., Mozetič, B. Sentiment of Emojis. PLoS ONE 10(12). 2015. [↑](#footnote-ref-9)
10. Liu, Bing. "Sentiment Analysis and Subjectivity." An chapter in Handbook of Natural Language Processing, Second Edition, (editors: N. Indurkhya and F. J. Damerau), 2010. [↑](#footnote-ref-10)
11. Maynard,, D., & Greenwood, M. A. (n.d.). “Who cares about sarcastic tweets? Investigating the impact of sarcasm on sentiment analysis” n.d. Web. 01, Dec. 2016. [↑](#footnote-ref-11)
12. Lu, X., Ai, W., Liu, X., Li, Q., Wang, N., Huang, G., and Mei, Q. 2016. Learning from the ubiquitous language: an empirical analysis of emoji usage of smartphone users. In Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '16). ACM, New York, NY, USA, 770-780. [↑](#footnote-ref-12)
13. *Ibid.* [↑](#footnote-ref-13)
14. Tigwell, G. W., and Flatla, D. R..“Oh that's what you meant!: reducing emoji misunderstanding.” In Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct (MobileHCI '16). 2016. ACM, New York, NY, USA, 859-866. [↑](#footnote-ref-14)
15. Novak. [↑](#footnote-ref-15)
16. Boia, M., Faltings, B., Musat, C. C., and Pu, P. 2013. A : ) is worth a thousand words: now people attach sentiment to emoticons and words in Tweets. In Proceedings of the 7th International Conference on Weblogs and Social Media, ICWSM2013. 345-350. [↑](#footnote-ref-16)
17. Foran, C. “The 'Never Clinton' Campaign. 5”, May 2016. Web. 01, Dec. 2016. [↑](#footnote-ref-17)
18. Yan, H., Sgueglia, K., & Walker, K. 'Make America White Again': Hate speech and crimes post-election. 29. Nov. 2016. Web. 01, Dec. 2016. [↑](#footnote-ref-18)
19. Shafer, J. “Donald Trump Talks Like a Third-Grader. Retrieved” 13, Aug. 2015. Web. 01, Dec. 2016. [↑](#footnote-ref-19)
20. Taboada, M., Brooke, J., Tofiloski, M., Voll, K., & Stede, M. Lexicon-Based Methods for Sentiment Analysis. Computational Linguistics, 37(2), 2011. 267-307. [↑](#footnote-ref-20)
21. Pang, B., Lee, L., and Vaithyanathan. S. “Thumbs up? Sentiment classification using machine learning techniques.” In Proceedings of the Conference on Empirical Methods in NLP, 2002. Philadelphia, PA. 79–86, [↑](#footnote-ref-21)
22. Turney, Peter. “Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews.” In Proceedings of 40th Meeting of the Association for Computational Linguistics, 2002. Philadelphia, PA. 417–424. [↑](#footnote-ref-22)
23. Nielsen, Finn Årup. "AFINN." *AFINN*. Informatics and Mathematical Modelling, Technical University of Denmark, Mar. 2011. Web. 28 Jan. 2017. [↑](#footnote-ref-23)
24. Nielsen, Finn Årup. "AFINN: A New Word List for Sentiment Analysis on Twitter." *Finn Årup Nielsen's Blog.* N.p., 25 Aug. 2015. Web. 28 Mar. 2017. [↑](#footnote-ref-24)
25. Nielsen, Finn Årup. "A New ANEW: Evaluation of a Word List for Sentiment Analysis in Microblogs." (n.d.): n. pag. Dtu Informatics, Technical University Of Denmark, Lyngby, Denmark. Web. 28 Mar. 2017. [↑](#footnote-ref-25)
26. *ibid* [↑](#footnote-ref-26)
27. Novak. [↑](#footnote-ref-27)
28. Rothenberg, Matthew. "Emojitracker: Realtime Emoji Use on Twitter." *Emojitracker: Realtime Emoji Use on Twitter*. N.p., n.d. Web. 28 Apr. 2017. [↑](#footnote-ref-28)
29. Laurenancona. "Laurenancona/twimoji." *GitHub*. N.p., 11 July 2015. Web. 28 Apr. 2017. [↑](#footnote-ref-29)
30. Prismoji. "PRISMOJI/emojis." *GitHub*. N.p., n.d. Web. 28 Apr. 2017. [↑](#footnote-ref-30)
31. Tufts, Chris. "Twimoji: Identifying Emoji in Tweets." *Mining the Details.* N.p., 24 July 2015. Web. 28 Apr. 2017. [↑](#footnote-ref-31)
32. Lowe, William. Benoit, Kenneth. Mikhaylov, Slava. and Laver, Michael. "Scaling Policy Preferences From Coded Political Texts." *Legislative Studies Quarterly* 26(1, 2) 2011. 123-155. [↑](#footnote-ref-32)
33. *Ibid.* [↑](#footnote-ref-33)
34. Silge, Julia, and David Robinson. *Text Mining with R*. N.p., 23 Apr. 2017. Web. 28 Apr. 2017. [↑](#footnote-ref-34)
35. Allison, Bill, Mira Rojanasakul, Brittany Harris, and Cedric Sam. "Tracking the 2016 Presidential Money Race." *Bloomberg.com.* Bloomberg, 09 Dec. 2016. Web. 28 Apr. 2017. [↑](#footnote-ref-35)
36. Duggan, M. “The Demographics of Social Media Users.” 19, Aug. 2015. Web. 02, Dec. 2016. [↑](#footnote-ref-36)
37. Smith, A. “Civic Engagement in the Digital Age.” 25, Apr. 2013. Web. 30, Nov. 2016. [↑](#footnote-ref-37)
38. Jack, R. E., Blais, C., Scheepers, C., Schyns, P. G., and Caldara, R. “Cultural confusions show that facial expressions are not universal.” *Current Biology* 19, 18. 2009. 1543-1548. [↑](#footnote-ref-38)
39. Lu, X., Ai, W., Liu, X., Li, Q., Wang, N., Huang, G., and Mei, Q. “Learning from the ubiquitous language: an empirical analysis of emoji usage of smartphone users.” *In Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '16).* ACM, New York, NY, USA. 2016. 770-780. [↑](#footnote-ref-39)
40. Park, J., Barash, V., Fink, C., and Cha, M.. 2013. “Emoticon style: interpreting differences in emoticons across cultures.” *In Proceedings of the Seventh International Conference on Weblogs and Social Media,* ICWSM 2013. 2013. [↑](#footnote-ref-40)
41. Lu. [↑](#footnote-ref-41)
42. *Ibid.* [↑](#footnote-ref-42)
43. Tigwell. [↑](#footnote-ref-43)
44. Maynard,, D., & Greenwood, M. A. “Who cares about sarcastic tweets? Investigating the impact of sarcasm on sentiment analysis.” n.d.. Web. 01, Dec. 2016. [↑](#footnote-ref-44)
45. Kleinman, Z. “Authorities 'use analytics tool that recognises sarcasm'” 2013. Web. 02, Dec. 2016. [↑](#footnote-ref-45)
46. Schott, B. “Ben Schott on the Strange New Vocabulary of the 2016 Election.” 2016. Web. 02, Dec. 2016. [↑](#footnote-ref-46)
47. Stack, L. “‘Braggadocious?’ In Trump and Clinton Debate, Words That Sent Viewers to the Dictionary.” 2016. Web. 02, Dec. 2016. [↑](#footnote-ref-47)
48. Brown, L. “Twitter - 30 Must-Know Twitter Abbreviations and Acronyms.” 2014. Web. 25, Nov. 2016. [↑](#footnote-ref-48)
49. Nielsen, Finn Årup. "A New ANEW: Evaluation of a Word List for Sentiment Analysis in Microblogs." (n.d.): n. pag. Dtu Informatics, Technical University Of Denmark, Lyngby, Denmark. Web. 28 Mar. 2017. [↑](#footnote-ref-49)
50. Lowe. [↑](#footnote-ref-50)