Exploring the Sentiment of Qatar World Cup Tweets

Humza Khan, Rebecca Larsen, and Shawn Stewart

EPPS6323 Knowledge Mining

Spring 2023

Dr. Karl Ho

**Abstract**

In this study, sentiment analysis is used to investigate how Twitter users feel about labor rights violations in Qatar in the years leading to the Qatar World Cup in 2022. A dataset including 3.5 million tweets was scraped for this study from Twitter, while the sample used for the study was smaller. For the purpose of conducting sentiment analysis, a vocabulary of roughly 6800 positive and negative English sentiment terms was used from the Word-Emotion Association vocabulary developed by the National Research Council of Canada. The study employs an automated technique to extract information regarding the attitudes, emotions, and opinions of Twitter users.

**Introduction**

Throughout the process of constructing the necessary infrastructure for the 2022 FIFA World Cup in Qatar, there have been a number of complaints of violations of workers' rights. It is estimated that Qatar relied on the labor of two million migrant workers in order to successfully stage the 2022 World Cup (Acharya. 2022). According to Amnesty International's 2019 report, both men and women, predominantly from Africa and Asia, worked in the construction, transportation, security, and restaurant industries.

The infringement of workers' rights can generally be categorized into three distinct categories. The first problem is the abuse of migrant workers. A significant portion of the workforce that was utilized in the construction of World Cup infrastructure came from the countries of Bangladesh, India, and Nepal, respectively. Reports indicate that these workers are regularly exploited in a variety of ways, including having their salary withheld, being forced to live in deplorable conditions, and having their passports taken, which prevents them from leaving the country or switching employment. The Kafala system comes in second. The Kafala system is used in a number of Gulf states. This system connects the legal rights of workers to those of their employers. Because they are unable to switch jobs without the approval of their current employer, they are susceptible to exploitation as a result of this situation. According to the claims, the personnel working in Qatar for the World Cup may have taken advantage of adopting this strategy. Third, there is the potentially dangerous working environment. As a result of the extreme temperatures in Qatar, multiple people have passed away there from heatstroke. Reports indicate that workers may have been expected to put in long hours without enough breaks or safety equipment, which may have been a contributing factor in the occurrence of accidents and injuries.

According to Amnesty International (2019), in 2017, the government passed a number of new legislative laws with the intention of assisting migrant workers. These bills included a statute for local workers, new labor dispute tribunals, and the establishment of a workers' assistance and insurance fund. However, doubts regarding the execution of these procedures and the overall treatment of personnel participating in developing World Cup infrastructure continue to be raised (Mohamed, 2022). The "International Labour Organization" is just one of the many organizations that have raised concerns regarding these issues.

A new dataset that was developed using data from the Twitter API will serve as the foundation for the research design. Text mining is the methodology that underpins this study (Silge & Robinson, 2022). The purpose of this study is to analyze, through the use of text and sentiment analysis, the shifts in opinion that have been stated in recent years on labor rights breaches in Qatar in advance of the FIFA World Cup 2022. The sentiment of tweets will be the dependent variable in this study. The years 2015 to 2022 have been designated as the time span for this study. The research will concentrate on determining how people's feelings shifted over the course of time leading up to the World Cup.

## Literature Review

Web scraping was used to collect data from Twitter, and sentiment analysis was used to evaluate the emotions that are conveyed in tweets (Silge and Robinson, 2022; Mohammad and Turney, 2011). A score was given to each tweet based on whether or not it expressed a positive or negative sentiment. Emotional experience is an inherent aspect of being human. Others maintain that the ways in which we communicate and the traditions we uphold significantly influence the emotions that we experience. Others contend that language and culture have a substantial influence on how we feel, despite the fact that some people say that facial expressions representing fundamental human emotions are universal throughout cultures, even those without contact (Ekman and Friesen, 2003; Ekman, 2005), and others contend that this is the case.

Psychologists have developed numerous theories to classify human emotions, including the distinction between basic and complex emotions as well as instinctual and cognitive emotions (Zajonc, 1984; Lazarus, 1984, 2000). Due to a lack of empirical evidence, however, the problem of defining emotions remains unresolved (Plutchik, 1985). According to the theories of Ekman (1992), Plutchik (1962, 1980, and 1994), and James (1884), there are categorizable fundamental human emotions. Plutchik proposes eight fundamental emotions, including happiness, trust, fear, surprise, sadness, disgust, anger, and anticipation. Ekman proposes six emotions, including joy, sadness, rage, fear, disgust, and astonishment.

This study's emotion lexicon, developed by Muhammad and Turney (2011), annotates words with Plutchik's eight fundamental emotions, which are arranged in a wheel-like pattern. Happiness, sadness, anger, and fear are positioned opposite one another on the emotional wheel. For instance, happiness is the opposite of sadness, and anger is the opposite of fear. Two of the primary emotions make up each of the four remaining emotions. Disgust, for instance, combines anger and fear, whereas confidence combines happiness and anticipation.

**Research Design and Methodology**

We depend on peer-reviewed text and sentiment analysis on the Qatar World Cup to inform our development of a methodology for keywords, hashtags, and analysis in relation to the World Cup (Dewi & Arianti, 2020; Luo et al., 2022; Patel & Passi, 2020). This allows us to design a methodology for keywords, hashtags, and analysis that is specific to the World Cup. Using the work of these scholars as a foundation, we will now concentrate on protecting worker rights.

**Data**

Using Twitter's Application Programming Interface (API), a new dataset was compiled as described Pak and Paroubek (2010). R, more specifically Rstudio, was used for both the data scraping, statistical analysis, and data visualization. R provided a platform where the accounts and packages were based on open source and were free (Ho, 2022). Quarto was used for a range of tasks, including creating documents with R Markdown and research reports.

For the purpose of data collection, the "academictwitteR" package was used. A wide variety of industries, such as customer relationship management, human-computer interface, information retrieval, text-to-speech systems, and literary analysis, make use of sentiment analysis. (Turney & Mohammed, 2011). In recent years, data from Twitter has been utilized in social science research on a regular basis (Dang-Xuan et al., 2013; Jungherr, 2016). For instance, policymakers utilize the data from Twitter as a vital resource in order to gain insights from the platform's analytics (Joseph et al., 2017). An additional illustration is provided by Parmelee and Bichard (2012), who study the reasons why Twitter users follow political personalities. Twitter is utilized for a variety of purposes, including ease of use, entertainment, self-expression, direction, information-seeking, and the value it adds to social interactions.

Web scraping was done to acquire data from Twitter (Silge and Robinson, 2022). Based on the emotional lexicon developed by Mohammed and Turney (2011), the data from Twitter was used to conduct a sentiment analysis in order to measure the sentiment included in the tweets.

Research was done using the “search_tweets()” function to make a list of hashtags and return the number of tweets for each hashtag. The eight hashtags with the greatest number of tweets were used. The hashtag limit was set by the “academictwitterR” package in this part of the preliminary research. The common hashtags identified and used to scrape data were "#boycottqatar2022," "#boycottqatar," "#worldcupqatar," "#worldcupqatar2022," "#qatarworldcup," and "#qatarworldcup2022." The data was scraped from 2015 to 2022. There were almost 3.5 million tweets in the dataset.

Silge and Robinson's (2017) approach to combining the NRC emotions and sentiment polarity was used with our dataset. Initially, the tweets were unnested, and stop words were removed. Additional stop words have been uncovered. The most frequent terms in the dataset include Qatar, FIFA World Cup, 2022, etc. Since it is already known that our research topic is the 2022 FIFA World Cup in Qatar, these do not aid in understanding what is happening in a given tweet or what topics are being discussed. These words were eliminated to determine how our analysis changes when these common words are eliminated. Additionally, we remove the hashtags that we used to collect the data (since they are present in every tweet). Note that we also took into account common misspellings of "Qatar" in these additional stop words (i.e., "Qatar").

**Sentiment Analysis**

For the purpose of this study, the Word-Emotion Association Lexicon developed by the National Research Council of Canada (Mohammad and Turney, 2011) will be referred to. The technique of finding the underlying emotional state conveyed by a string of words, in this case those found on Twitter, is referred to as "Twitter sentiment analysis. A customer's attitudes, feelings, and views can be gleaned in a meaningful way through the use of an automated process known as sentiment analysis, which can be performed using a sentiment analysis tool.

The NRC Emotion Lexicon is a collection of English words and their associations with two sentiments (negative and positive) and eight fundamental emotions (disgust, joy, sadness, surprise, trust, anticipation, anger, and fear). The two sentiments are negative and positive, while the eight basic emotions are anger, fear, anticipation, trust, and anticipation. The NRC Emotion Lexicon makes use of association scores in its methodology. In terms of emotions, it is given a score of either 0 (indicating that it is not associated) or 1 (indicating that it is associated). When it comes to emotions, it is given a score of "not associated," "weakly associated," "moderately associated," or "strongly associated."

**Sample Full Sentiment Analysis**

The NRC Word-Emotion Association Lexicon by Mohammad & Turney (2022) was used. After encountering difficulties with the NRC full sentiment analysis timing out or taking more than 10 hours to complete, we approached the NRC Lexicon by combining the NRC emotions with our dataset and exploring each emotion separately. We conducted a small test to determine whether this method would execute faster than the Syuzhet package functions. This functioned and ran effectively. Next, we expanded upon this test by transforming this code into a function that we call for each emotion in the NRC Lexicon. The NRC loaded with this code originated from tidytext and consists of 13872 observations representing 10 distinct sentiments (8 emotions and 2 polarity indicators).

**Word Clouds**

We created word clouds using two different approaches. The first applied Silge & Robinson's (2017) method to the individual word data frame. The second relied on Raghav Bali, Dipanjan Sarkar, and Tushar Sharma's (2017) book to further explore full tweet word clouds and clustering. These authors use tm_map to clean the data, and then plot a word cloud.

**Emotions by Top Words**

Here we create lists of top words by polarity and emotion. Then, we generate bar graphs of the most frequent words associated with each emotion. We assign colors to each emotion bar plot based on Nijdam's (2009) research, "Mapping emotion to color," so that colors are applied based on empirical assessments of color and emotional response.

**Full tweets rather than individual words**

To explore the sentiment of each full tweet, we aggregated tweets by tweet IDs. This provided us with a count of emotion-related words per tweet, from which we were able to determine the sentiment of the tweet. For example, if a tweet contained two words associated with happiness and eight words associated with sorrow, it would be classified as sad.

**Hierarchical Clustering**

Raghav Bali, Dipanjan Sarkar, and Tushar Sharma (2017) use the Syuzhet package to get sentiment scores and plot them. This was still very slow, as we experienced with other functions from the Syuzhet package. Therefore, we shifted to relying on Bali, Shakar, and Sharma (2017)'s approach for hierarchical clustering. This is an unsupervised learning approach to clustering data.

We recombined the tidy data to put the words of each tweet back together, minus the stop words to show how words show up in clusters together.

**Exploring labor-related keywords by count and over time**

We finished our exploratory analysis by considering when tweets were created. We transformed the time variable into a recognized numeric form with date and time of post. We created a histogram to understand the frequency of when tweets were posted in our sample. Then we focused on concerns over labor rights. We selected key labor-related words that were informed by our earlier emotion bar plots, transformed them to lowercase, and created dataframes and total counts. We plotted these labor rights words over time to understand the time variations in labor rights related tweets over our sample period.

**Inferential Analysis**

After we completed our unsupervised approaches and explored the results, we used these to inform some testable hypotheses to determine what drove polarity of tweets. We relied on Logistic Regression given the lack of a linear relationship assumption. Our dependent variable is a two-class binary variable (positive or negative tweet sentiment), therefore this classification method is most appropriate (James et al., 2021).

**Modeling**

Using logistic regression for supervised learning, we will adopt a classification-based approach to modeling our data. To begin modeling our data, we will need a data frame containing all of the relevant variables. Our tweets_tidy_sum dataframe contains both the NRC information and the tweet text, and we joined it with the original tweets dataframe to obtain additional metrics.

**Splitting Train/Test Data**

To begin modeling, we split our data into a training group and a testing group. We used a 70/30 split, with 70% of the data used to fit the model and 30% used for testing purposes. It is important to only use the training data subset to fit the model, including variable selection; otherwise, we will not get an accurate estimate of the test error (James et al. 2021). We assign cases to the test or training set using the sample() function, which draws a random sample. It is important that the cases be assigned randomly; otherwise, we risk introducing bias into our model.

**The Best Subset Selection**

Best subset selection is one option for selecting a R2 of predictors for our model. It generates a number of potential models, which can then be evaluated based on AIC, BIC, and adjusted R^2. It becomes computationally intensive, or even infeasible, with a large number of possible predictors. However, for our model using only fifteen predictors, it was still manageable.

James et al. (2021) discuss the advantages of forward or backward stepwise selection methods when $p$ is large. When $p$ is very large, the best subset selection method may lead to overfitting, and in those cases, a stepwise approach may be better. However, in this case with only $p = 15$, we decided to go with the best subset selection.

We started with a logistic regression model containing all possible predictors from our data set – sadness, anger, fear, disgust, trust, surprise, joy, anticipation, year, retweets, likes, impressions, replies, quotes, and impressions – and regressed that on polarity. Then, using the leaps package and regsubsets() function, we performed the best subset selection. For the function, we set the maximum number of predictors to 15, the same number of predictors in the model. This created a summary with 15 possible models. The first model contains only one predictor, the second contains two, and so on, with the best subset selection returning the best model for each number of predictors (James et al. 2021).

From this point, we need to determine the ideal number of predictors. To do that, we graph information available from the regsubsets() summary. We looked at RSS, adjusted R^2, AIC, and BIC. All summaries showed model improvement up to eleven predictors, after which benefits of adding more predictors dropped off.

**K-fold Cross Validation**

Now that we have selected a model, we want to estimate its prediction accuracy using k-fold cross validation. K-fold cross validation can be used to estimate testing error, and it can also be used for the purpose of model selection (James et al. 2021). In this case, we are most interested in the test error rate. Using the caret package, we set up a training control and run the train() function. The results tell us that our selected model has an 82% accuracy rate.

**Exploratory Results**

**Word Clouds**

Word clouds in text analysis provide an easy way to present huge volumes of data. Words are placed in a cloud-like structure to visually display text data in word clouds, with the size of each word denoting its frequency (Cooshna-Naik, 2022). They are a straightforward way to see the most frequent terms in a text, and they are frequently used in big data sets to find themes, subjects, or keywords. Word clouds have a number of benefits, including the ability to summarize data on a subject, pinpoint key passages in a document, and determine which areas or subjects to focus on when conducting further statistical inquiry.

**Figure 1.**

***Word Clouds***

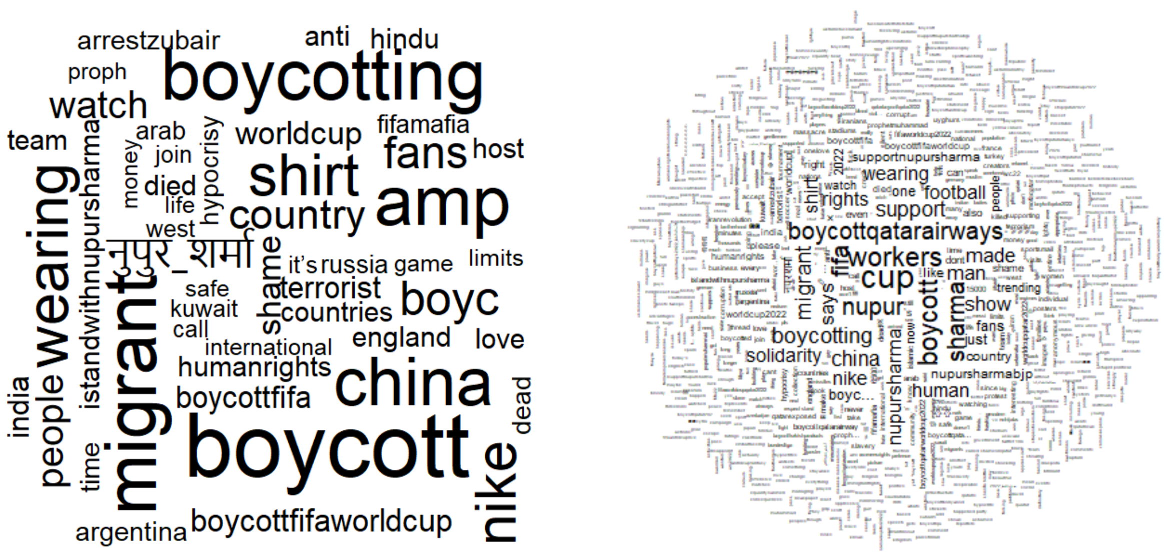



Figure 1 identifies the words that most frequently appear in the tweets. Important text identified by the word cloud include “boycott, migrant, boycottfifaworldcup, fifamafia, humanrights, nike, india, life, died, and hypocrisy”. We found the first approach to be more interpretable.

## Sentiment Analysis

A polarity sentiment analysis reveals that there are more negative associated words than positive words in our data set (see Table 1). To better understand the context of the positive/negative discussions, we turned to keyword counts overall and then by emotion. We found the bar charts to more clearly reveal the top used words, which built on the initial exploration from the word clouds above (Figure 1). Figure 2 shows the top 30 most frequently used words in our dataset. Boycott is the most common, with workers, nupur, sharma, and boycottairaways following closely. We observe several areas of frustration being discussed in relation to boycotts, particularly concerning Indian politician Sharma Nupur and airline controversies. In relation to our initial interest in labor rights surrounding the World Cup, we interpret “workers” and “migrants” to be connected to concerns over labor rights, given that migrant workers predominantly built the arenas and human rights reports revealed abuse, neglect, deaths, and non-payment (Amnesty International, 2022). When isolating negative sentiment, the top tweeted word remains boycott. Words we interpret as being associated with labor rights are slavery and massacre. Variations of terror and corruption also have high frequencies but are too general to connect directly to labor rights violations.

**Table 1.** ***Overall Positive vs. Negative Sentiment by Count***

| **Negative** | **Positive** |
|---|---|
| 3316 | 2308 |

Figures 3 to 11 display the top words for each emotion in the NRC Emotion Lexicon. As mentioned above, we are applying Nijdam’s (2009) findings on mapping emotional response to color. We first discuss the negatively associated emotions and top words (Figures 4 to 8). First, looking at the results for anger, the words terrorist, money, terrorism, slavery, massacre, and hate are included. Slavery is directly connected to our labor rights issue. We assume money is also related but not as consistently. This is based on a manual exploration of tweets using the word money. They often referred to workers not being paid but other tweets varied greatly in their context. Similarly, when examining fear associated words (Figure 5), we again see slavery directly relating to labor rights and shame sometimes being associated but varying greatly in context. “Watch" could be used in a concerning and fearful way - such as watch out, but in the context of the World Cup it is likely that “watch” simply refers to *watching* the football games. We are uncertain that the word “watch” is properly interpreted by the analysis. Given a manual inspection, we observe instances where tweets are referencing watching the football matches without any detection of a fear-based use of the word watch. Next, slavery and massacre occur in top words associated with sadness. Figure 9 displays two top used words associated with surprise. We first listed this with our positive words but after seeing money and terrorism, and confirming with a manual inspection, we realized that surprise is being used in a predominantly negative manner. Beyond the scope of labor rights, we see terrorism and its variations showing up as a top word for each of the negative emotions. This is clearly driving negative boycott tweets outside the scope of our variable of interest. We understand that several areas of concern are informing negative tweets about boycotting the Qatar World Cup.

**Figure 2**

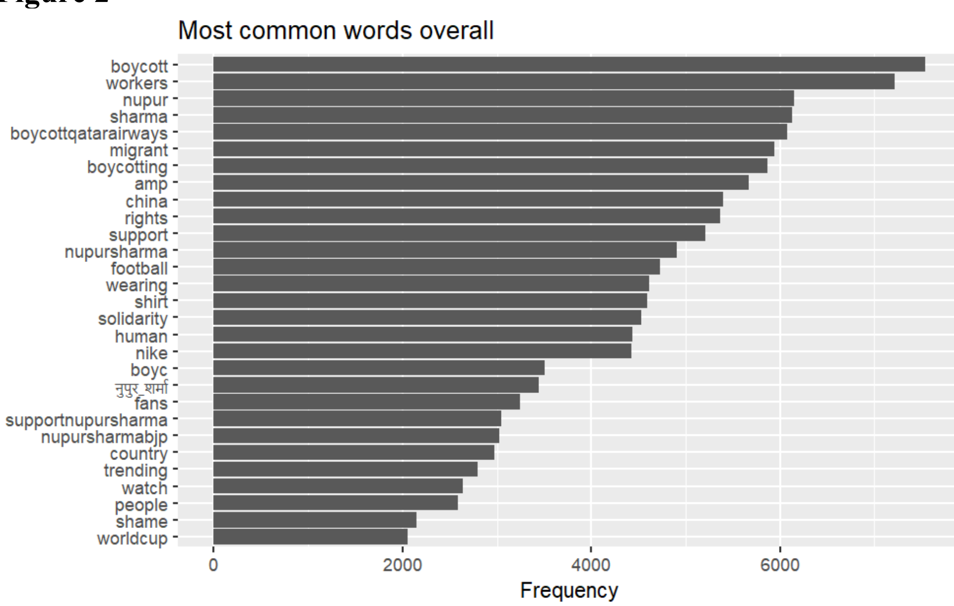


**Figure 3**

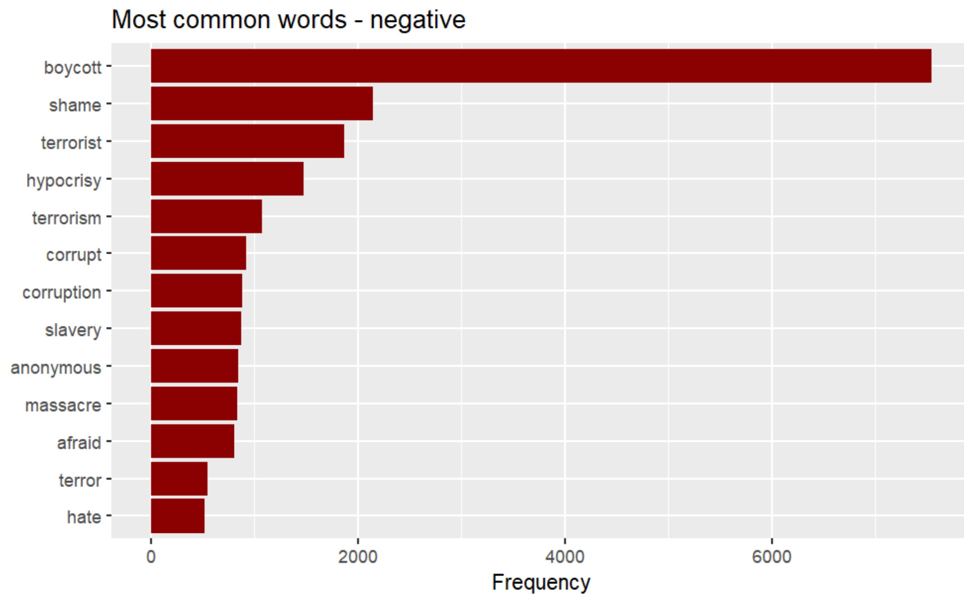



**Figure 4**

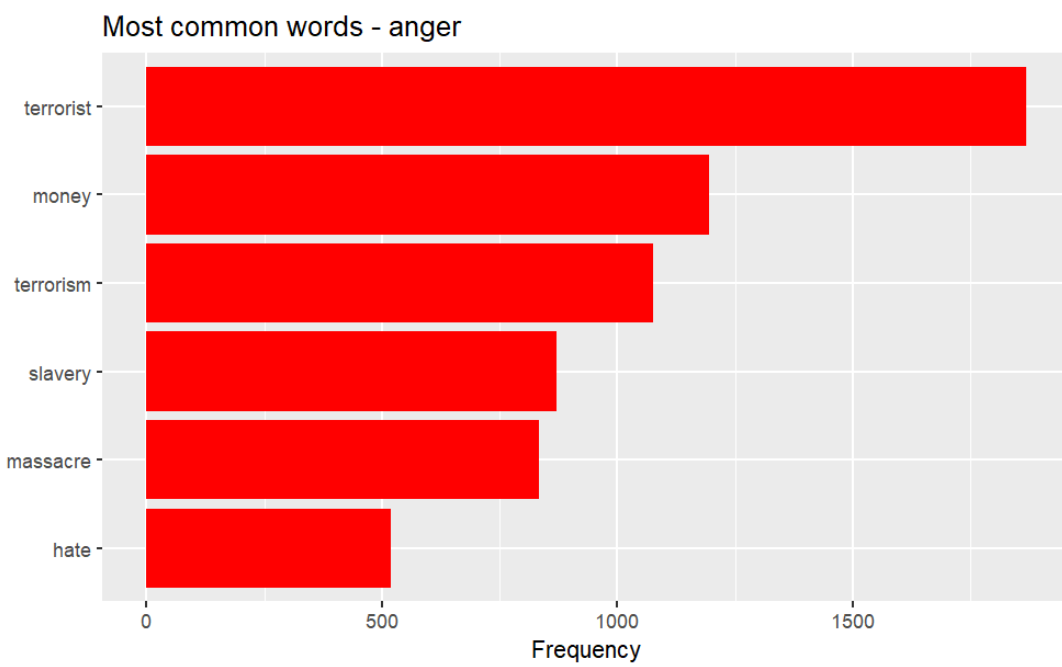


**Figure 5**

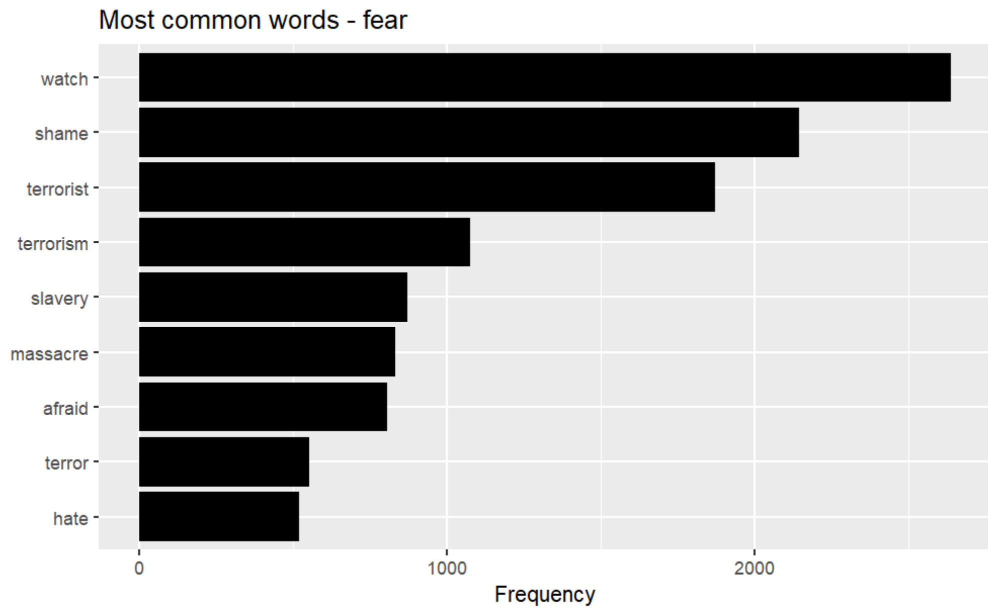



**Figure 6**

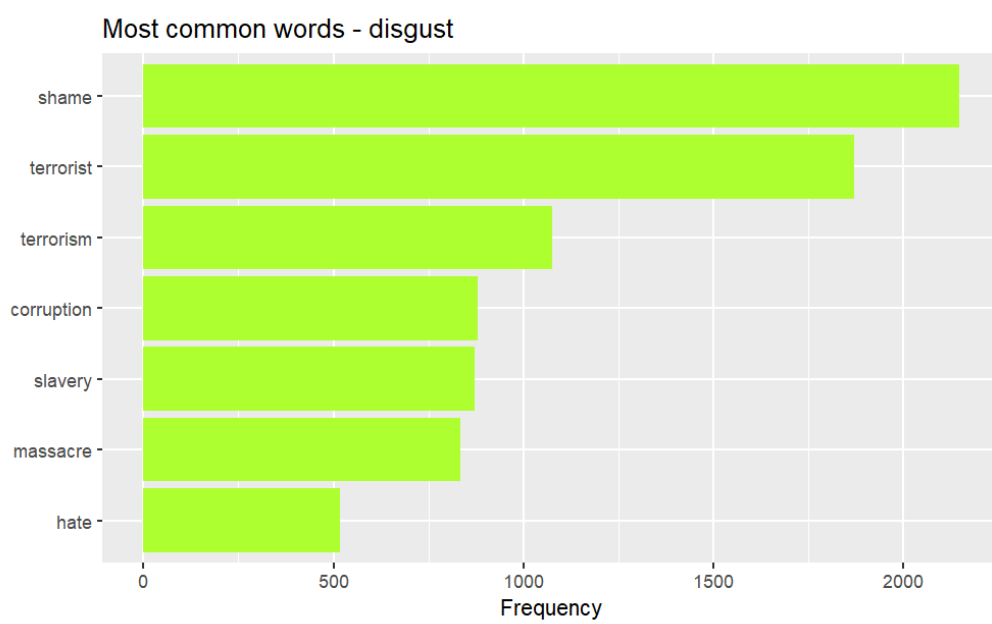


**Figure 7**

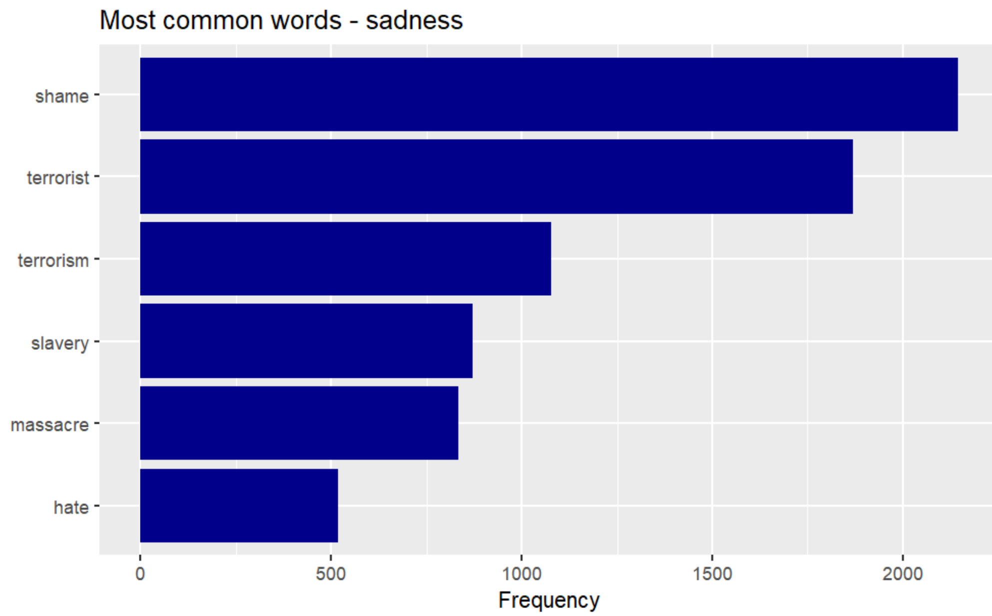
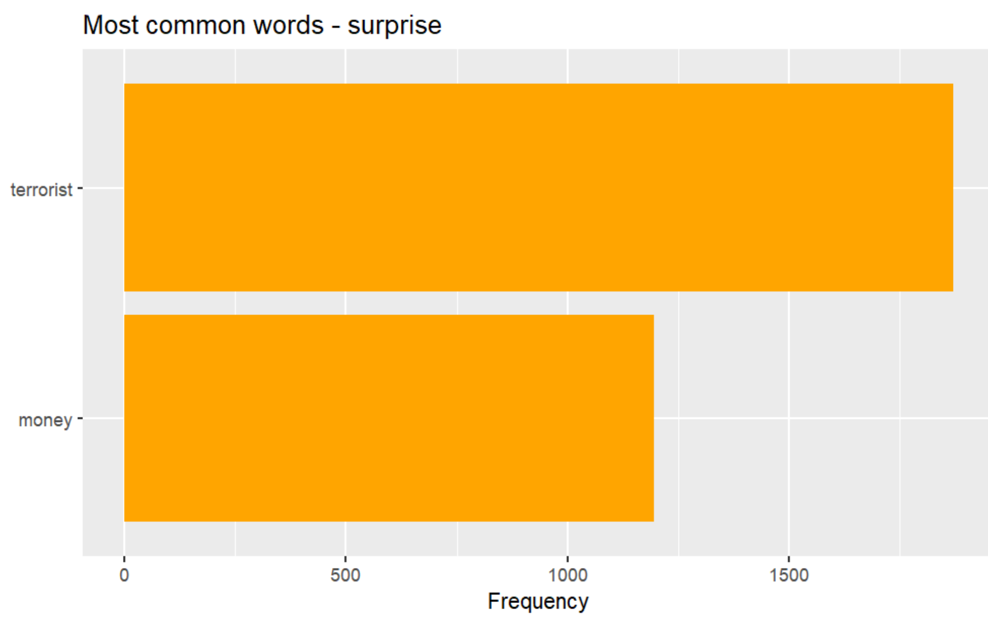


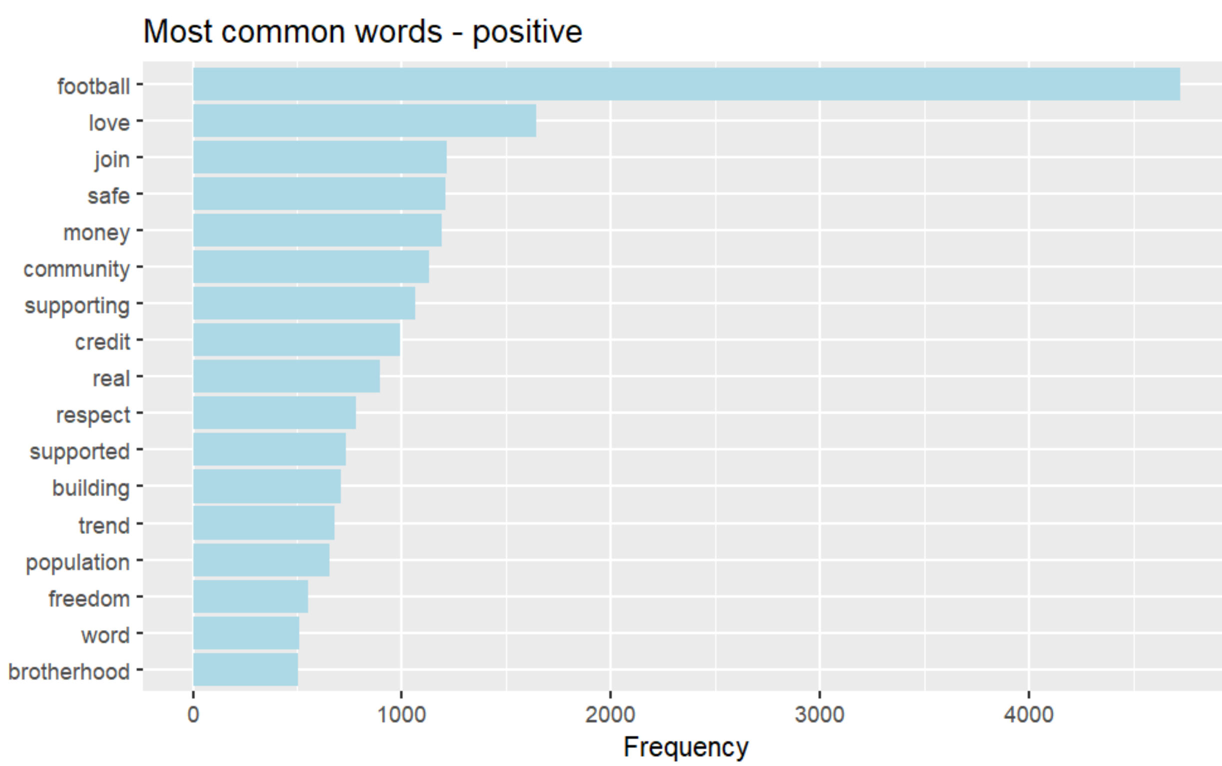

**Figure 8**




Though positive tweets were less frequent in our dataset, there were still a large quantity of positively identified tweets, and their key words represented notably different discussions. Figure 9 shows the frequency of the top 30 most commonly used words in positively associated tweets.

Here we see football being the most tweeted word. Like the word “watch” in the top words, we are uncertain that football is being properly interpreted here. A manual inspection reveals positive uses but also frequent negative uses about the footballers participating in the midst of labor rights abuses and the brands that support them. We are not certain that all positive sentiment readings of football in Figure 9 are accurate. It is also a top used word related to anticipation and joy. Interestingly, money shows up in both positive and negative sentiments, but otherwise all other words are unique to positive sentiments.

**Figure 9**




Solidarity is a top trust-associated tweet (Figure 10). There are tweets that suggest trust in activists, trust of workers, and trust of specific footballers. This could therefore be interpreted positively or negatively, depending on who trust is being connected to in the tweet. Freedom is another high frequency word that is present on the positive, trust, and joy figures (Figure 11, 12, & 13). Inspecting freedom tweets, we observe people celebrating freedom, and also calling out a lack of freedom. Again, the context can be positively or negatively associated with the World Cup.

**Figure 10**

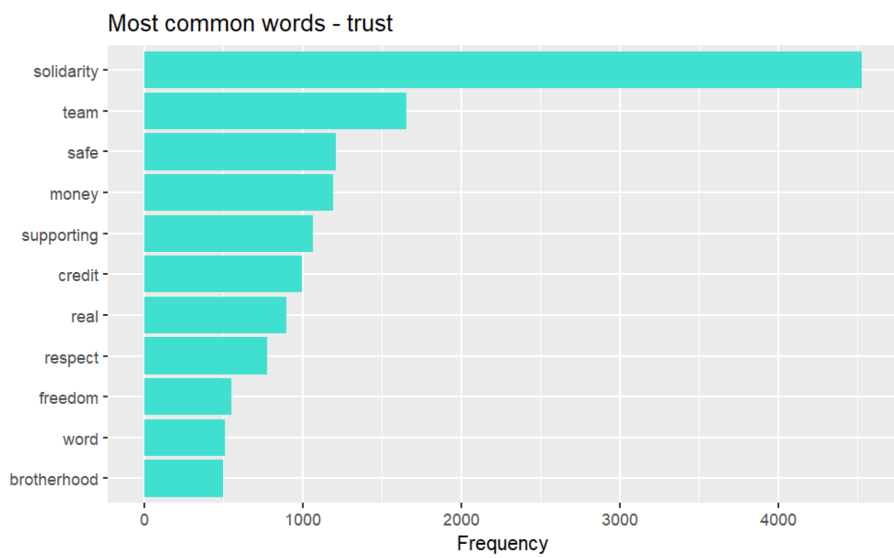



**Figure 11**

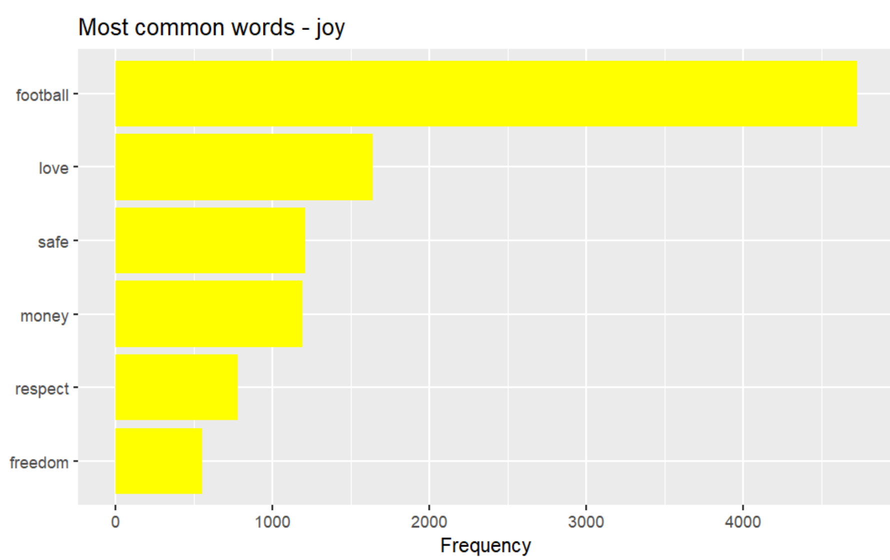


**Figure 12**

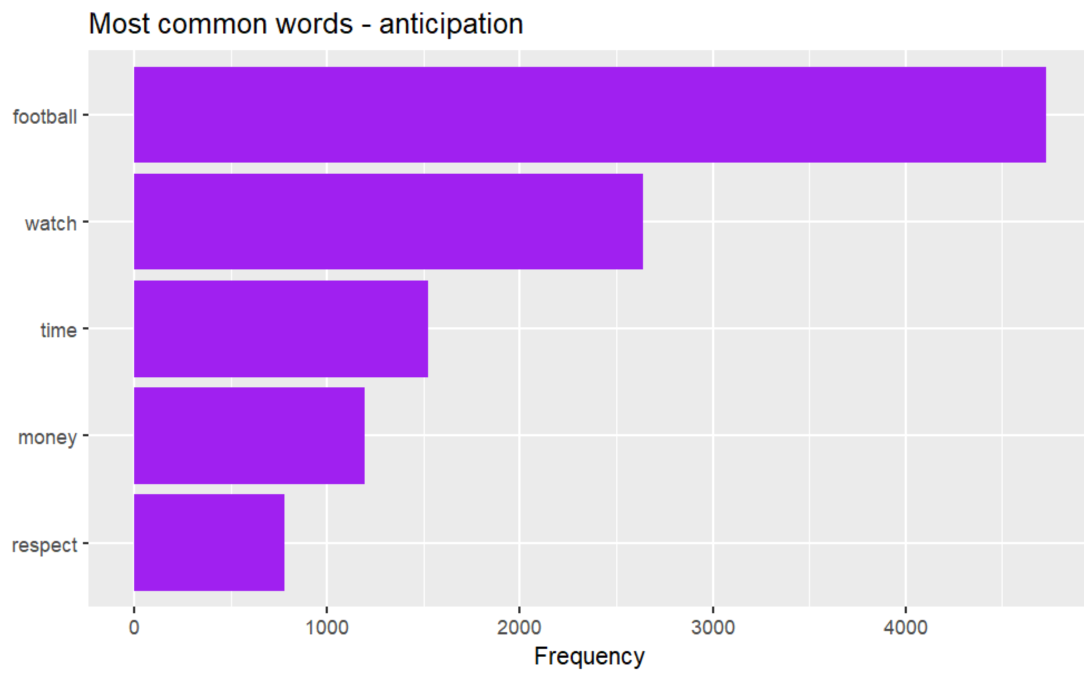
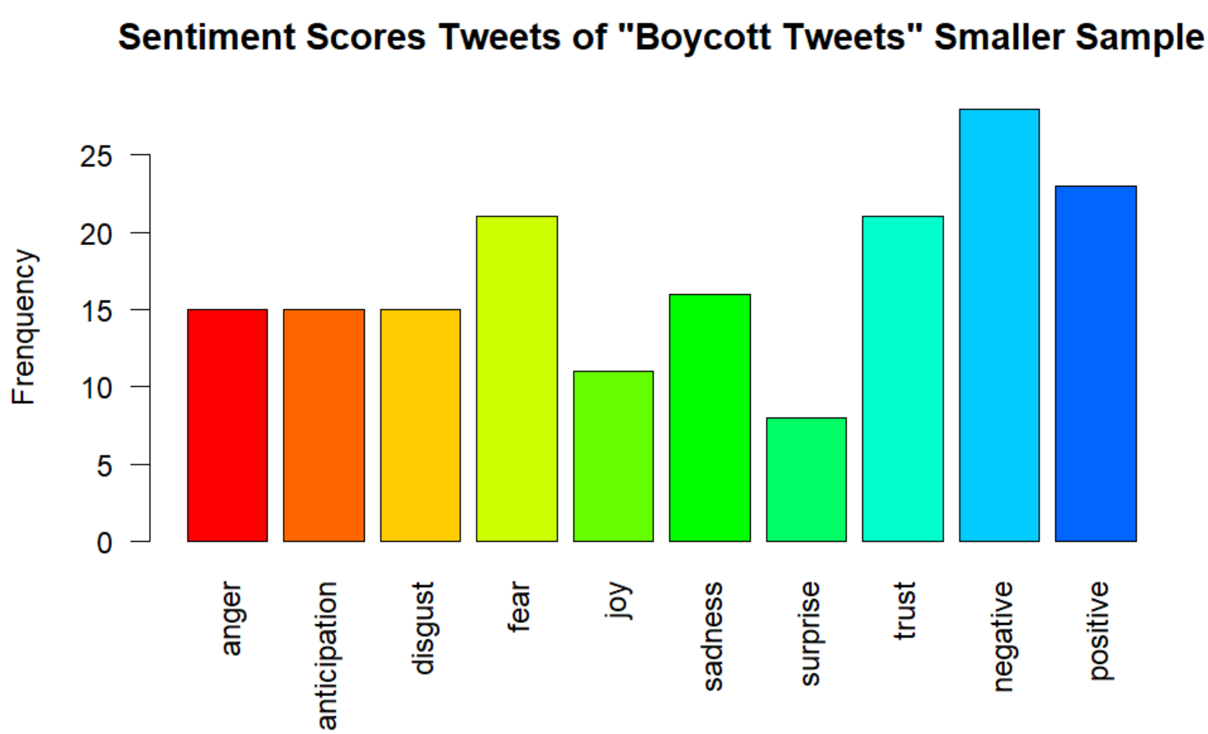



**Figure 13**



Beyond the individual exploration of key words for each main emotion, Figure 13 presents the results of a full sentiment analysis on a sample of our tweets. This approach considers the full tweet instead of individual words, and charts them comparatively to one another. We again see higher frequencies of negative tweets compared to positive ones, with fear and trust being the emotions most frequently connected to the tweets, followed by sadness. Joy and surprise have the lowest frequency, while anger, anticipation, and disgust each have the same frequency.

**Hierarchical Clustering**

**Figure 14**

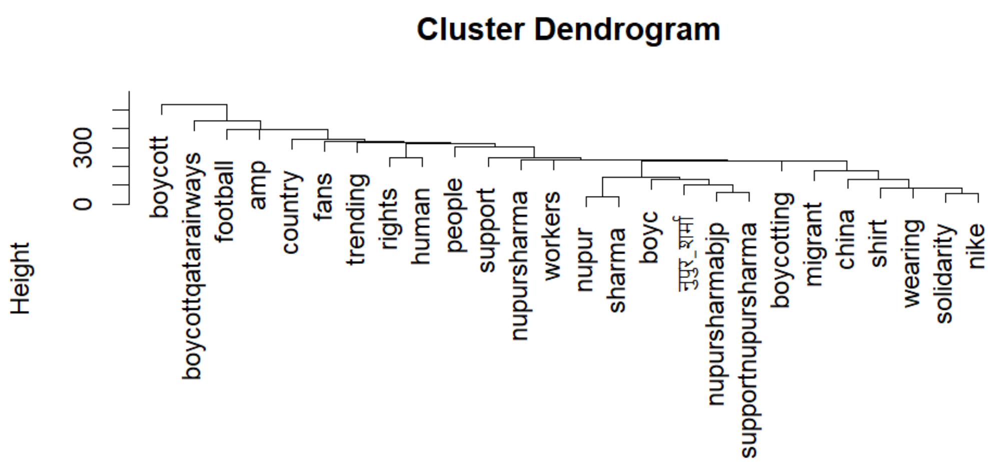



Figure 14 presents a dendrogram to illustrate how words are clustered together and associated. We created this to help us understand not just sentiments but also what kinds of words are being discussed together to better understand context. Though there is not a definitive insight from this figure, we do notice a few patterns. Beginning at the right of the graph, we observe right, shirt, wearing, solidarity, and Nike together. This begins to offer context that we could develop further analysis of to understand if, perhaps, people were discussing wearing Nike shirts in protest. Similarly, with Nupur Sharma, we observe different variations of her name, boycott, and support being connected. Though this is outside of the labor rights context, there is more that could be analyzed to capture the level of support she received given that this former BJP spokesperson was being criticized by Qatar and tweets were responding to the controversy. We also note that humans and rights are connected. We wonder if, even though labor rights were heavily scrutinized in relation to the Qatar World Cup, people are not thinking about rights as separate types but connecting these types of violations into a larger discussion of human rights.

**Time**

We end our exploratory analysis with a set of histograms to explore when tweets were made. Our data set selected tweets from 2015 to 2022. Figure 15 reveals, predictably, that the majority of the tweets took place during the World Cup and leading up to it. However, there was also a cluster in 2015. This was after the previous Brazil World Cup in 2014, and we perceive it to reflect controversy around the early planning of the Qatar World Cup. We then pulled out specific labor-related keywords that showed up in our previous analysis to understand when those tweets occurred. We saw that same surge of tweets in 2015 for boycott tweets that mentioned “labor”, “death”, and “slavery” (Figure 16), but “construction”, “work”, and “rights” (Figures 16 & 17) did not have the same spike in 2015. Future research could use this first step to understand if the 2017 labor rights legislation in Qatar had an impact on the concern over labor rights leading up to the World Cup.

**Figure 15**

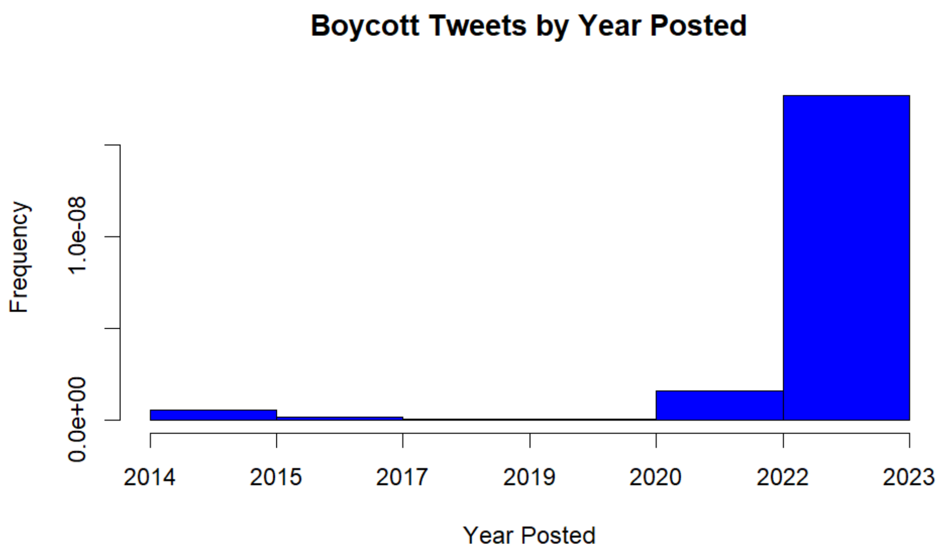



**Figure 16**

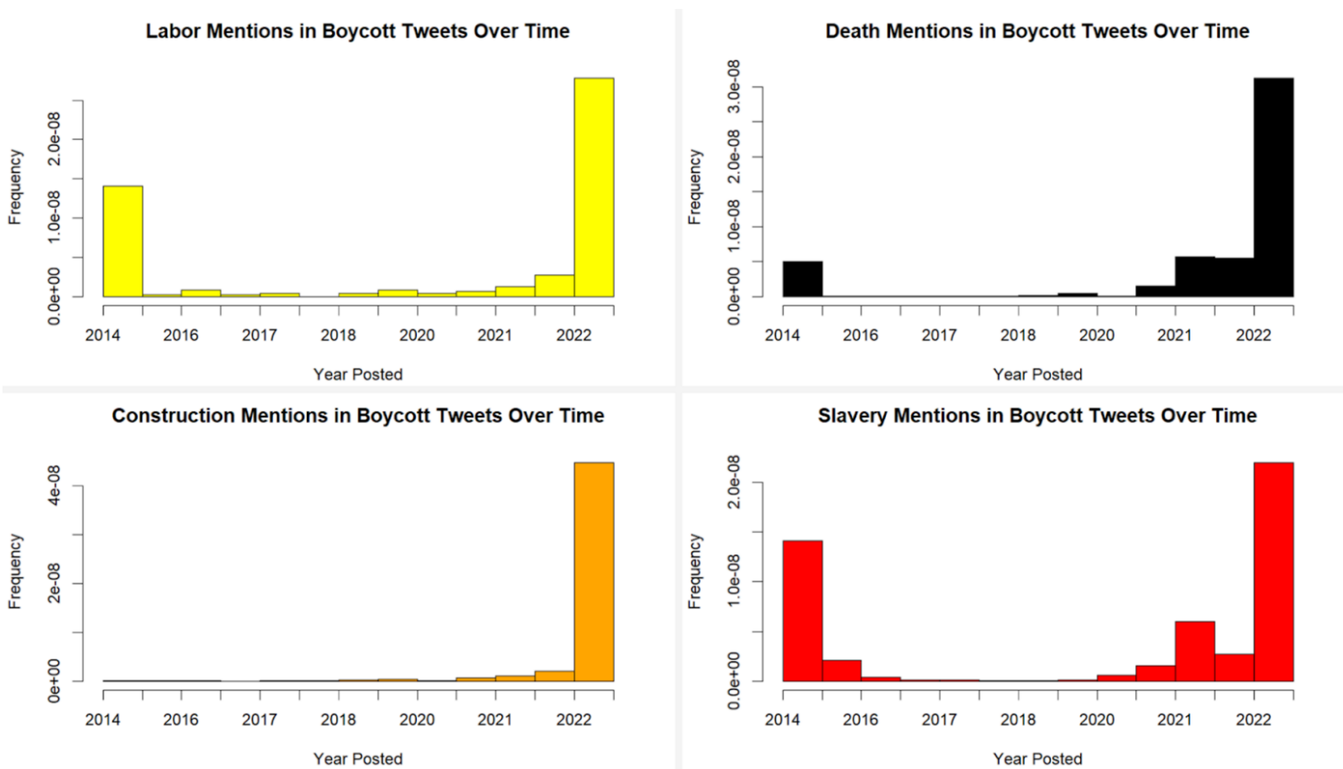
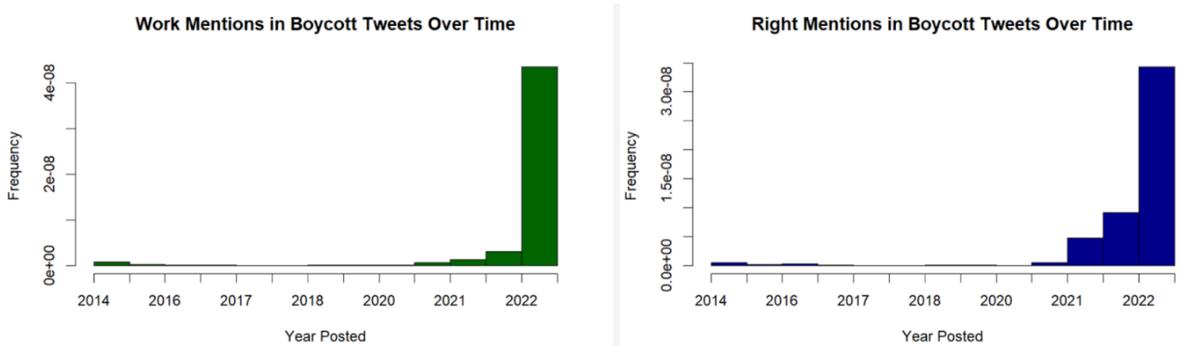


**Figure 17**




## Inferential Results

We began by identifying the dependent variable and all possible predictors in our dataset. Our dependent variable is the “polarity” of the tweet. Any tweets that had a higher count of negative words than positive words were given a value of 1 (negative). Neutral or positive tweets are represented as 0. Including all NRC emotions and years, along with other Twitter metrics, gives us a total of 22 possible predictors. Table 2 shows the results of a simple logistic regression using all 22 predictors.

Table 2 shows a large number of predictors are statistically significant. However, with such a large number of possible predictors it is very likely that there is multicollinearity among the predictors, especially since some constructs are closely related (such as the NRC emotions, which are naturally grouped into negative and positive designations).

**Table 2.**

All Possible Predictors on Polarity

| | *Dependent variable:* |
|---|---|
| | polarity |
| sadness | 0.799*** |
| | (0.031) |
| anger | 0.330*** |
| | (0.027) |
| fear | 0.973*** |
| | (0.024) |
| disgust | 1.058*** |
| | (0.030) |
| trust | −1.107*** |
| | (0.025) |
| surprise | −0.690*** |
| | (0.040) |
| joy | −1.144*** |
| | (0.033) |
| anticipation | −0.343*** |
| | (0.026) |
| 2016 | −0.884*** |
| | (0.149) |
| 2017 | −0.534** |
| | (0.250) |
| 2018 | −1.196*** |
| | (0.265) |
| 2019 | −0.238 |
| | (0.224) |
| 2020 | −3.219*** |
| | (0.107) |
| 2021 | −1.325*** |
| 2022 | −0.520*** |
| | (0.059) |
| 2023 | −0.730*** |
| | (0.261) |
| Retweets | −0.0004*** |
| | (0.00002) |
| Likes | −0.0002 |
| | (0.0003) |
| Impressions | −0.0001 |
| | (0.0001) |
| Replies | 0.011*** |
| | (0.004) |
| Quotes | −0.009 |
| | (0.007) |
| Constant | −0.682*** |
| | (0.058) |
| Observations | 75,477 |
| Log Likelihood | −30,032.260 |
| Akaike Inf. Crit. | 60,108.520 |
| *Note:* | *p<0.1; **p<0.05; ***p<0.01 |

**Table 3**

Correlation Matrix of Independent Variables

| | fear | sadness | anger | disgust | trust | surprise | joy | anticipation | retweets | likes | impressions | replies | quotes |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **fear** | 1 | | | | | | | | | | | | |
| **sadness** | 0.729825 | 1 | | | | | | | | | | | |
| **anger** | 0.678633 | 0.71171 | 1 | | | | | | | | | | |
| **disgust** | 0.669104 | 0.759066 | 0.738999 | 1 | | | | | | | | | |
| **trust** | 0.177223 | 0.21018 | 0.251389 | 0.180386 | 1 | | | | | | | | |
| **surprise** | 0.353356 | 0.40795 | 0.4589 | 0.370028 | 0.357892 | 1 | | | | | | | |
| **joy** | 0.142003 | 0.194063 | 0.220064 | 0.165254 | 0.49372 | 0.387125 | 1 | | | | | | |
| **anticipation** | 0.276969 | 0.188407 | 0.214958 | 0.155066 | 0.374881 | 0.382533 | 0.660868 | 1 | | | | | |
| **retweets** | -0.17571 | -0.16355 | -0.16784 | -0.15717 | 0.123149 | -0.12067 | -0.1722 | -0.18161283 | 1 | | | | |
| **likes** | 0.003654 | 0.002499 | 0.002798 | 0.001552 | 0.002249 | 0.00334 | 0.004147 | 0.003252873 | 0.008393 | 1 | | | |
| **impressions** | 0.004455 | 0.001303 | 0.003973 | 0.007801 | 0.004949 | 0.008629 | 0.011356 | 0.010385944 | -0.00645 | 0.007611 | 1 | | |
| **replies** | 0.009579 | 0.007323 | 0.009407 | 0.00652 | 0.006546 | 0.014579 | 0.011983 | 0.011263966 | -0.00341 | 0.713457 | 0.005976775 | 1 | |
| **quotes** | 0.001763 | 0.001281 | 0.001589 | 0.002423 | -0.00092 | 0.000555 | 0.000108 | 0.000546613 | 0.001718 | 0.466848 | 0.001487316 | 0.624155 | 1 |

We generated a correlation matrix to take a look at the correlations between our predictors. What this shows us is that many of our predictors are correlated with each other, and the negative emotion words are highly correlated with each other. When considering the reliability of our models, we need to be mindful of overfitting.

After reviewing the simple logistic regression, we used best subset selection to identify which variables to include in our final model. We used the leaps package in R to run the best subset selection, and from the results, we reviewed the RSS, adjusted $R^2$, the AIC ($C_p$), and BIC results. As is typical, adding variables to the model improves it substantially at first. Then, as we continue adding variables, we see less and less improvement. Visually inspecting the charts in Figure 19, we see that model improvements dramatically decrease after adding six variables. Using R, we are able to notate the lowest point on each chart, which is consistent across all graphs. Based on the charts, we decided to use the 11-predictor model for our final model, although we could also have justified a more parsimonious model.

**Figure 18**

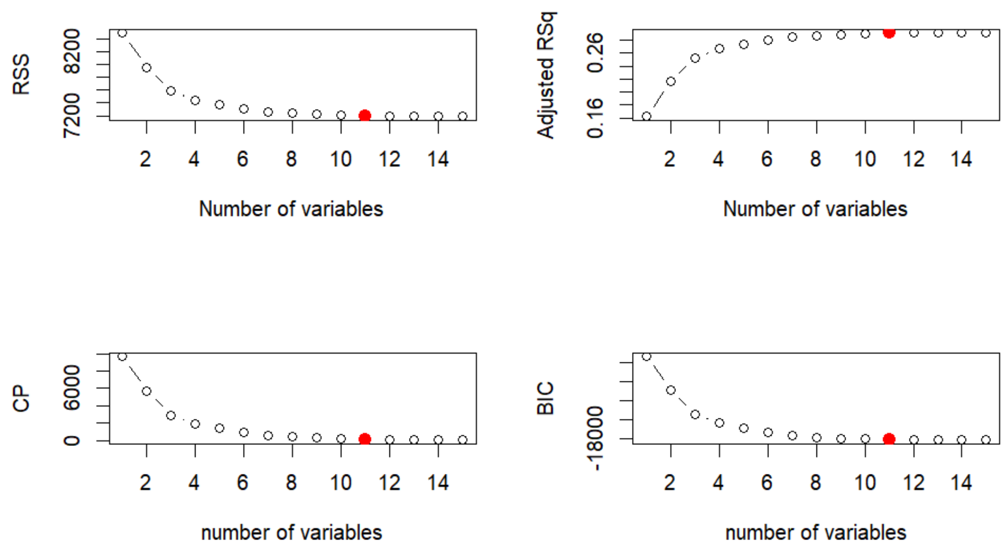



Table 4 shows the results of our final logistic regression model, using the top 11 predictors identified through best subset selection. We notice that the model includes three negative emotions, which were highly correlated with each other. Only one Twitter metric, number of retweets, made it into this final model. While this model reports a number of statistically significant coefficients, without better accounting for the effects of multicollinearity, we should be careful about drawing conclusions. One way to account for multicollinearity is through interaction terms, which are not included in our model here. It is possible that an even better model could be found by using interaction terms among the emotion predictors and the year predictors.

**Table 4**

Best Predictors on Polarity

| | *Dependent variable:* |
|---|---|
| | polarity |
| sadness | 0.905*** |
| | (0.035) |
| fear | 0.896*** |
| | (0.026) |
| disgust | 1.182*** |
| | (0.033) |
| trust | −1.034*** |
| | (0.029) |
| surprise | −0.593*** |
| | (0.046) |
| joy | −1.378*** |
| | (0.035) |
| 2020 | −2.749*** |
| | (0.120) |
| 2021 | −1.035*** |
| | (0.085) |
| 2022 | −0.301*** |
| | (0.059) |
| retweet | −0.0004*** |
| | (0.00002) |
| Constant | −0.915*** |
| | (0.058) |
| Observations | 52,833 |
| Log Likelihood | −21,242.850 |
| Akaike Inf. Crit. | 42,507.710 |
| *Note:* | *p<0.1; **p<0.05; ***p<0.01 |

After creating our model, we took one final step to validate it using k-fold cross validation. We performed the cross-validation using the caret package in R. We trained the model using the train() function in conjunction with trainControl(), then checked the accuracy of our model. The k-fold cross validation indicated that our model is 82% accurate. However, this could be an overestimation given the correlated predictors.

## Limitations and Challenges

There can be sample bias (Rivers & Vu, 2013). The study cannot guarantee that the sentiments presented in the tweets are representative of the sentiments of all the people from Qatar who expressed their sentiments on labor rights.

According to Gao, Abel, and Houben (2012), there is a possibility for measurement inaccuracy due to an inability to differentiate effectively between bot accounts and multiple-account human users who have access to several accounts. Due to the widespread presence of confounding factors, that cannot be properly managed or measured, the validity and reliability of the data are constrained. In particular, the inability to discriminate effectively between bots and human users who have several accounts may induce measurement bias, which has the potential to distort the results and limit the generalizability of the findings. As a result, extreme caution is required when interpreting the results of the study, and additional research needs to be carried out in order to overcome this restriction.

According to Fisher (1993), the validity of indirect inquiry might be put in jeopardy when social desirability bias is present. Due to the social desirability bias, some Twitter users might not convey their genuine feelings about a labor rights violation, which could have an effect on the validity of the study's findings.

Another limitation is that only participants who had access to the internet and a Twitter account were eligible. These people are more likely to be from socioeconomically affluent backgrounds than a sample that includes everyone involved.

Although the NRC Emotion Lexicon is a useful tool for studying Twitter data, its application may not always take into account the nuanced emotional expressions that are present in such data. It's also critical to keep in mind that context, irony, and sarcasm all have the potential to affect how accurate an analysis is. As a result, it might be more fruitful for future research to examine Twitter data using a multifaceted strategy that includes human annotation, machine learning algorithms, and context-specific result interpretation.

Even though sentiment analysis and logistic regression have the potential to analyze the topic of sentiment regarding labor rights violations in Qatar, it is important to keep in mind the limitations of the techniques used to gather data when interpreting the results of the research.

**Future Research**

Future research could include a larger date range to capture 2010 to 2022. This time period would capture the announcement of Qatar as host of the World Cup, capture the 2014 World Cup in Brazil to see if discussions of Qatar interacted with discussions during the 2014 World Cup, and understand if the 2017 legislation had a meaningful impact on sentiment of labor rights in Qatar.

Building on this, we could test the key words identified by our word cloud and our key words by emotion to see if labor-related keywords are predictive of negative tweets. Similarly, we could also expand the dependent variable to look at what drives the polarity of tweets.

**Conclusion**

In this report, we used unsupervised learning techniques to investigate the sentiment and context of boycott related tweets leading up to the Qatar World Cup in 2022. We were interested in understanding if labor rights were driving boycott tweets. We found some evidence to support this, but overall, we found a variety of emotions and keywords driving boycott tweets. After using word clouds and sentiment analysis to explore the data, we narrowed in on a dependent variable: the polarity of tweets. After testing for the optimal model, we determined that sadness, fear, and disgust were significantly increasing the likelihood of negative tweets. In the future, we would analyze a larger time span and consider other dependent variables that more directly connect to labor rights.

**References**

Acharya, S.P. (2022) "'Our Dreams Never Came True.' These Men Helped Build Qatar's World Cup, Now They Are Struggling to Survive". CNN. https://www.cnn.com/2022/11/17/football/qatar-2022-world-cup-migrant-workers-human-rights-spt-intl/index.html

Amnesty International (2019). "Reality Check: Migrant Workers' Rights in Qatar". https://www.amnesty.org/en/latest/campaigns/2019/02/reality-check-migrant-workers-rights-with-two-years-to-qatar-2022-world-cup/

Amnesty International. (2022). "FIFA: TIME TO COMPENSATE MIGRANT WORKERS IN QATAR". https://www.amnesty.org/en/latest/campaigns/2022/05/fifa-time-to-compensate-migrant-workers-in-qatar/

Bali, R., Sarkar, D., & Sharma, T. (2017). Learning Social Media Analytics with R. Packt Publishing.

Chaudhuri. (2019). Visual and Text Sentiment Analysis through Hierarchical Deep Learning Networks. Springer Singapore. https://doi.org/10.1007/978-981-13-7474-6

Cooshna-Naik, D. (2022). Exploring the Use of Tweets and Word Clouds as Strategies in Educational Research. *Journal for Learning for Development. Vol. 9*, No. 1, pp. 89-103

Dang-Xuan, L., Stieglitz, S., Wladarsch, J., Neuberger, C. (2013). An investigation of influentials and the role of sentiment in political communication on twitter during election periods. https://doi.org/10.4324/9781315680439-27

Dewi, S., & Arianto, D. B. (2023). Twitter Sentiment Analysis Towards Qatar as Host of the 2022 World Cup Using Textblob. Journal of Social Research, 2(2), 443-455.

Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion, 6*(3), 169–200.

Ekman, P., & Friesen, W. V. (2003). Unmasking the Face: A Guide to Recognizing Emotions From Facial Expressions. Malor Books.

Ekman, P. (2005). *Emotion in the Human Face*. Oxford University Press.

ElHabr, R. O. T. (2018). Converting nested JSON to a tidy data frame with R | R-bloggers. R-bloggers. https://www.r-bloggers.com/2018/10/converting-nested-json-to-a-tidy-data-frame-with-r/

Fisher, R. J. (1993). Social desirability bias and the validity of indirect questioning. Journal of Consumer Research, 20(2), 303-315. https://doi.org/10.1086/209351

Gao, Q., Abel, F., & Houben, G. J. (2012). A comparative study of microblogging behavior between Chinese and English users. In Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work (pp. 55-64). ACM. https://doi.org/10.1145/2145204.2145214

Ho, Karl (2022). Data Programming with R. https://datageneration.org/dataprogrammingwithr/

James, Gareth, Daniela Witten, Trevor Hastie and Robert Tibshirani. (2021). Introduction to Statistical Learning with Applications in R, New York, NY: Springer.

James, W. (1884). What is an emotion? Mind, 9, 188–205. https://www.jstor.org/stable/2246769

Joseph, N., Grover, P., Rao, P.K., Ilavarasan, P.V. (2017). Deep Analyzing Public Conversations: Insights from Twitter Analytics for Policy Makers. Digital Nations – Smart Cities, Innovation, and Sustainability. Lecture Notes in Computer Science. https://doi.org/10.1007/978-3-319-68557-1_22

Jungherr, A. (2016). Twitter use in election campaigns: A systematic literature review. Journal of Information Technology & Politics, 13(1), 72-91. https://doi.org/10.1080/19331681.2015.1132401

Lazarus, R. S. (1984). On the primacy of cognition. American Psychologist, 39(2), 124–129.

Luo, Y., Choi, J., Andon, S.P., Benton, B., & Green, K. (2022). Protests and calls for boycotting the Qatar World Cup 2022 Spark Online Discussion and Action Away from the Soccer Pitch. Montclair State University Digital Commons. https://digitalcommons.montclair.edu/scom-facpubs?utm_source=digitalcommons.montclair.edu%2Fscom-facpubs%2F32&utm_medium=PDF&utm_campaign=PDFCoverPages

Mohammad, S. and Turney, P. (2011). NRC Word-Emotion Association Lexicon (aka EmoLex). 2011 National Research Council Canada (NRC).

https://saifmohammad.com/WebPages/NRC-Emotion-Lexicon.htm

Mendon, Dutta, P., Behl, A., & Lessmann, S. (2021). A Hybrid Approach of Machine Learning and Lexicons to Sentiment Analysis: Enhanced Insights from Twitter Data of Natural Disasters. *Information Systems Frontiers, 23*(5), 1145–1168. https://doi.org/10.1007/s10796-021-10107-x

Mohamed, Noor. (2022). World Cup 2022: What has changed for migrant workers in Qatar? International Labour Organization. https://www.ilo.org/infostories/en-GB/Stories/Country-Focus/world-cup-qatar#landing

Parmelee, J.H., & Bichard, S.L. (2012). Politics and the Twitter Revolution: How Tweets Influence the Relationship between Political Leaders and the Public. Lexington Books. https://books.google.com/books?hl=en&lr=&id=ZcVxGdegAJ8C&oi=fnd&pg=PR5&ots=TN1GOOTLo-&sig=2vTVVB7Y0AkFsWhtwtLUrFl1WA0#v=onepage&q&f=false

Patel, R. Passi, K. (2020). Sentiment analysis on Twitter Data of World Cup Soccer Tournament Using Machine Learning. https://doi.org/10.3390/iot1020014

Periscopic. (2022, February 12). Cozy Collecting, Part 2: Working with Nested JSON in R. Medium. https://medium.com/@Periscopic/cozy-collecting-part-2-5e717588e37b

Plutchik, R. (1985). On emotion: The chicken-and-egg problem revisited. Motivation and Emotion, 9(2), 197–200.

Silge, J. and Robinson, D. Text Mining with R: A Tidy Approach. https://www.tidytextmining.com/