Twitter Scraping

Using twitterR package to search for the 100 thousand latest english tweet containing "cryptocurrency"

```
crypto_twitter <- searchTwitter("cryptocurrency",n=100000,lang="en")
crypto_twitter_df <- twListToDF(crypto_twitter)                                    # Convert to data frame
tweet_words <- crypto_twitter_df %>% select(id, text) %>% unnest_tokens(word,text)   # Unnest bulk of words

tweet_words %>% count(word,sort=T) %>% slice(1:20) %>%
        ggplot(aes(x = reorder(word, n, function(n) -n), y = n)) +
        geom_bar(stat = "identity") + theme(axis.text.x = element_text(angle = 60, hjust = 1)) + xlab("")
```

Create a list of stop words: a list of words that are not worth including

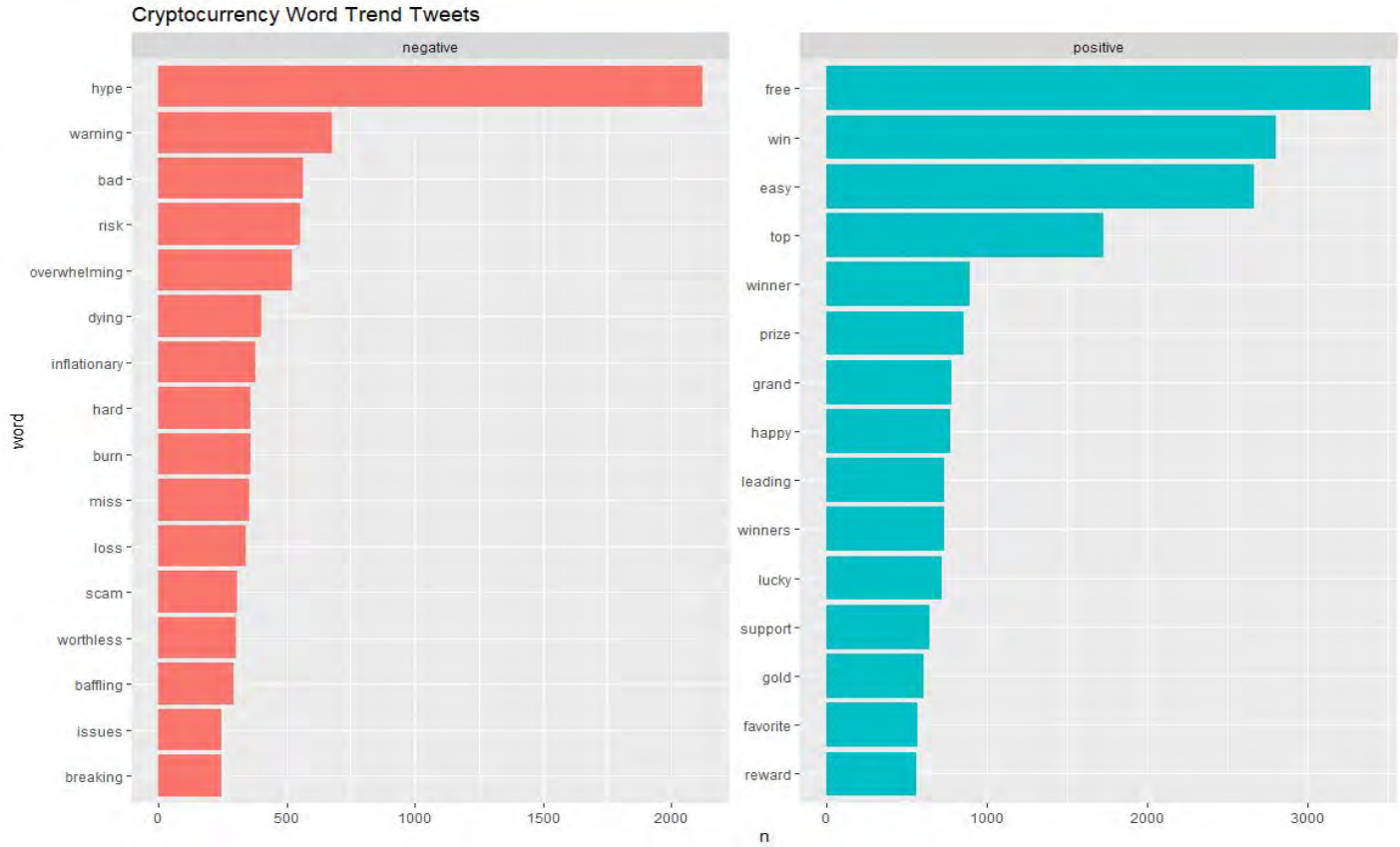
```
my_stop_words <- stop_words %>% select(-lexicon) %>%
        bind_rows(data.frame(word = c("a","and","in","is","https","for","of","on","partridge","rt","t.co",
                                      "this","that","trent","trentpartridge","with","y","you")))
tweet_words_interesting <- tweet_words %>% anti_join(my_stop_words)
tweet_words_interesting %>% count(word,sort = TRUE)                    # Give the counts of the top frequently appeared words
```

*Adjusting stop words list based on the outcome of the **previous NRC analysis** (these are the additional words that I want to remove from the dataset which is not inside the default stopword list command, because **every topic has its unique dictionary for sentiment**)*

```
my_stop_words <- tibble(word = c("1","2","3","4","amp","cap","cash","credit","exchange","follow","giveaway",
                                 "money","save","share","trade"),lexicon ="twitter" )
new_stop_list <- stop_words %>% bind_rows(my_stop_words)              #updated new stop word list
tweet_final <- tweet_words_interesting %>% anti_join(new_stop_list)       #remove new stopwords from dataframe
```

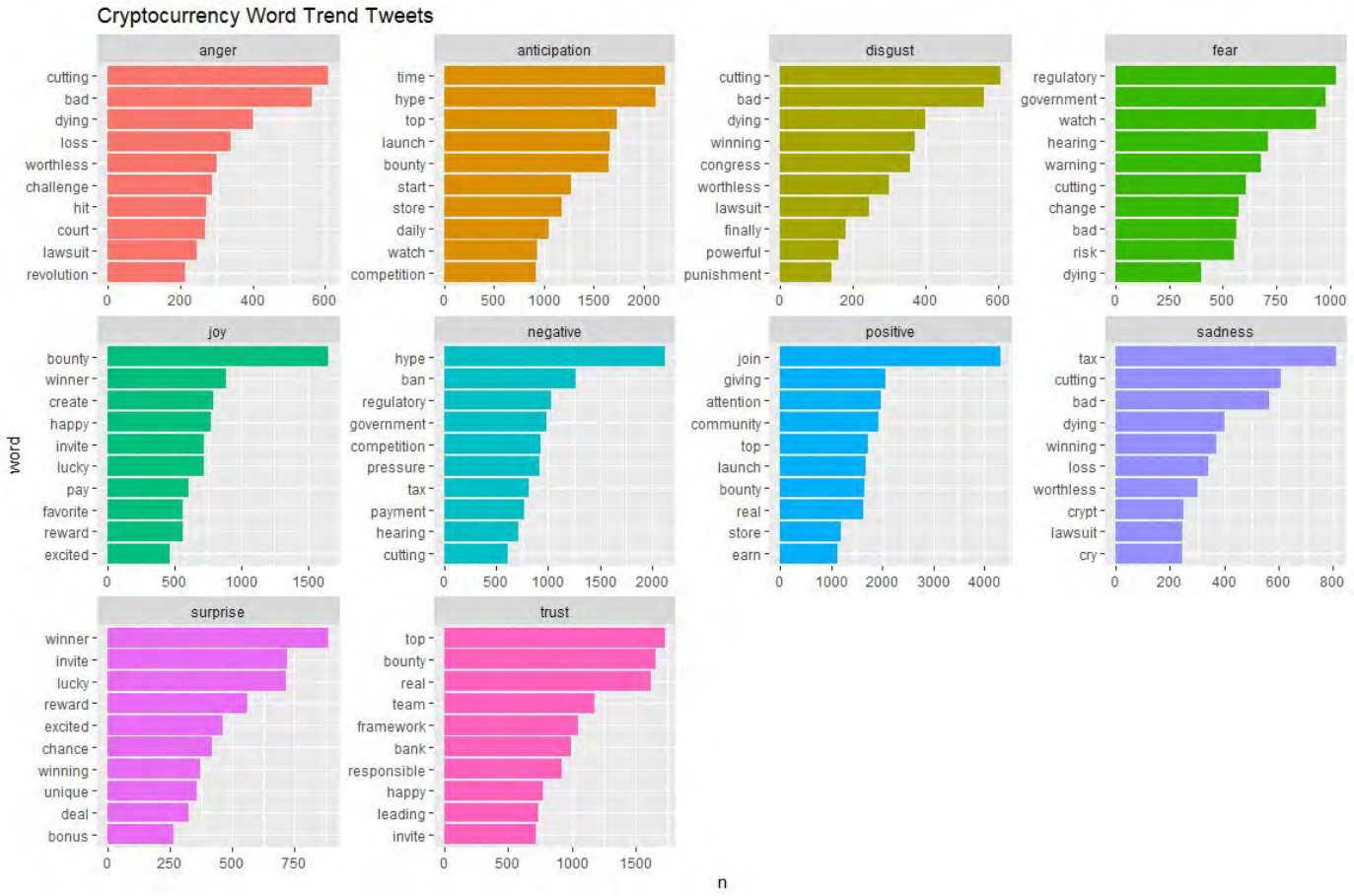
Bing Sentiment Analysis

```
crypto_sentiment1 <- tweet_final %>% inner_join(get_sentiments("bing")) %>% count(word, sentiment)
topwords <- crypto_sentiment1 %>% group_by(sentiment) %>%
        top_n(15) %>% ungroup() %>%   mutate(word = reorder(word, n))
ggplot(topwords, aes(word, n, fill= sentiment)) +
        geom_col(show.legend = FALSE) + facet_wrap(~sentiment, scales = "free") +
        labs(title= "Cryptocurrency Word Trend Tweets") + coord_flip()
```



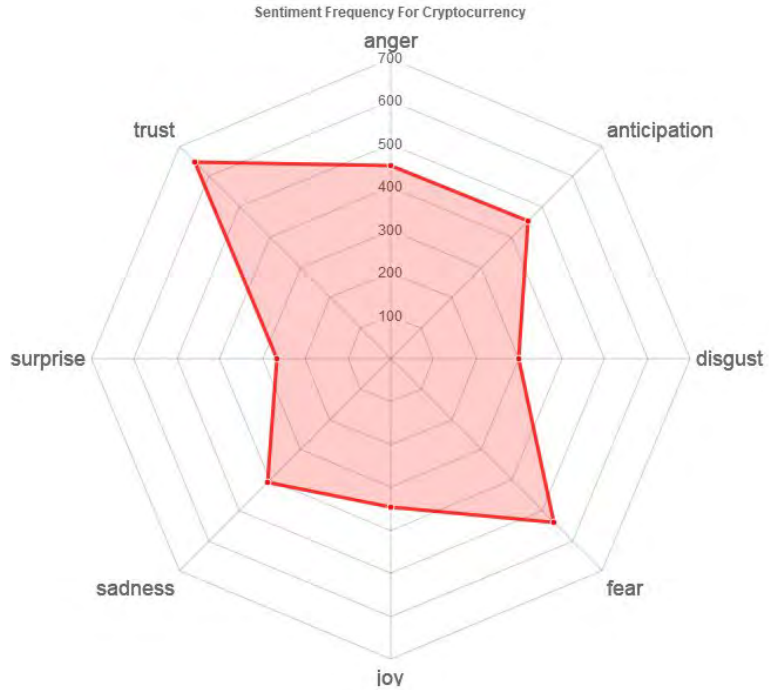
Cryptocurrency - NRC Sentiment Analysis

```
crypto_sentiment2 <- tweet_final %>% inner_join(get_sentiments("nrc")) %>% count(word, sentiment)
topwords <- crypto_sentiment2 %>%  group_by(sentiment) %>%
        top_n(10) %>% ungroup() %>%  mutate(word = reorder(word, n))
ggplot(topwords, aes(word, n, fill= sentiment)) +
        geom_col(show.legend = FALSE) + facet_wrap(~sentiment, scales = "free") +
        labs(title= "Cryptocurrency Word Trend Tweets") + coord_flip()
```



NRC Radar Chart

```
nrc_radar_chart <- crypto_sentiment2 %>%  filter(!grepl("positive|negative", sentiment))
word_tally <- nrc_radar_chart %>%  group_by(sentiment) %>%  tally()
chartJSRadar(scores= word_tally, showLegend = FALSE, main= "Sentiment Frequency For Cryptocurrency")
```



From the radar graph, we can infer the dominating sentiment is trust anticipation, fear and sadness. However, trust and fear are 2 almost opposite sentiments, so we must be careful with the conclusion we find in this analysis.

My next step is to look into what words count as trust and fear, and based on our "cryptocurrency" topic, add stop words into our stop word list, and hopefully get a more intuitive result out of the radar graph. I also want to suggest that the time frame on which the tweets have been tweeted can also have a critical impact on the results I get. Also by comparing "cryptocurrency" to similar topics like "bitcoin" and "libracoin" might draw totally different results.

What's the sentiment difference between January vs May vs June about "Cryptocurrency"?

Nope. Twitter won't let me.

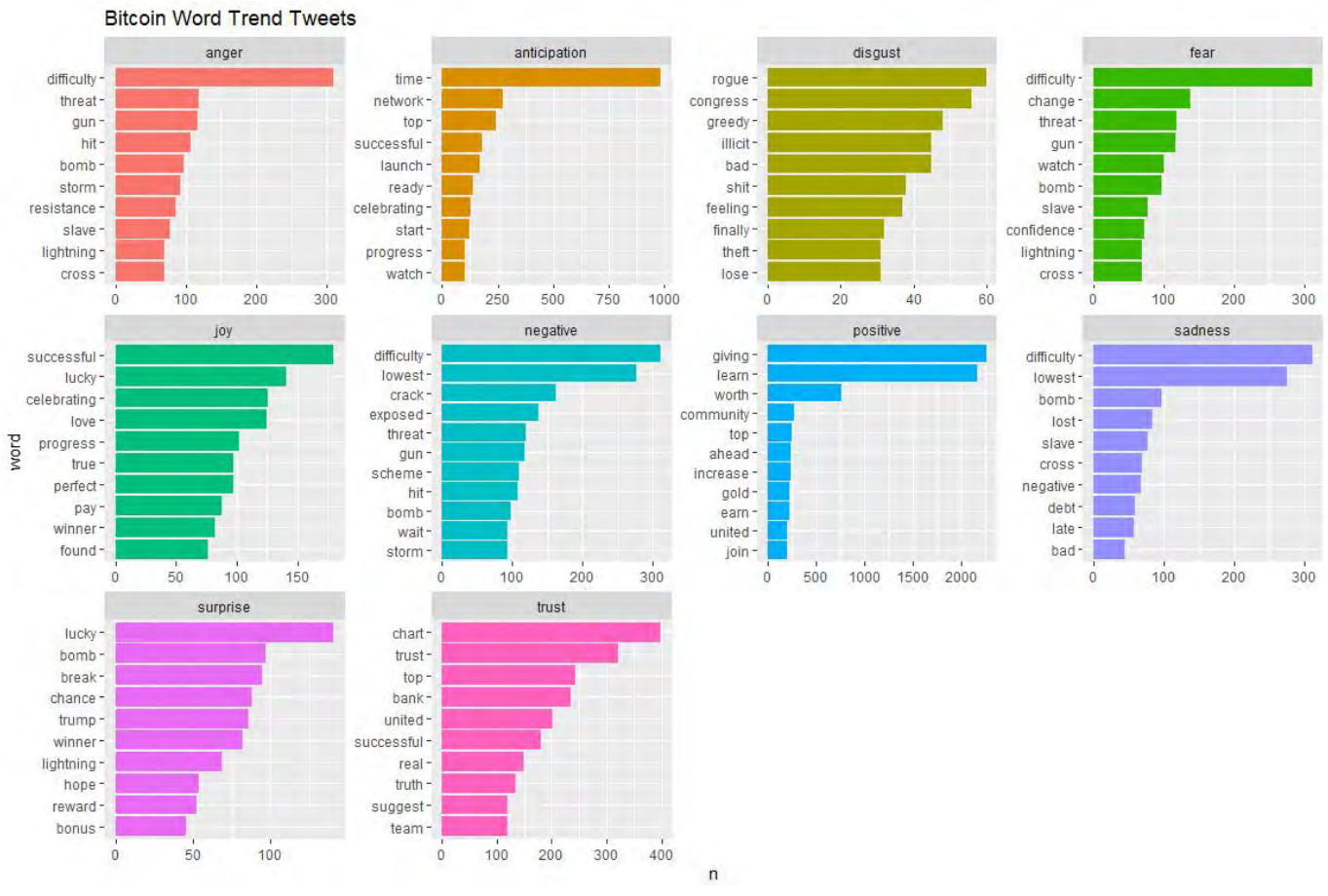
What about Bitcoin?

```
btc_twitter <- searchTwitter("bitcoin",n=20000,lang="en")
btc_twitter_df <- twListToDF(crypto_twitter)
tweet_words <- btc_twitter_df %>% select(id, text) %>% unnest_tokens(word,text)
my_stop_words <- stop_words %>% select(-lexicon) %>%
        bind_rows(data.frame(word = c("a","and","in","is","https","for","of","on","partridge","rt","t.co",
                                      "this","that","trent","trentpartridge","with","y","you")))
tweet_words_interesting <- tweet_words %>% anti_join(my_stop_words)

my_stop_words <- tibble(word = c("1","2","3","4","amp","cap","cash","credit","exchange","follow","giveaway",
                                 "money","save","share","trade"),lexicon ="twitter" )
new_stop_list <- stop_words %>% bind_rows(my_stop_words)
tweet_final <- tweet_words_interesting %>% anti_join(new_stop_list)
```

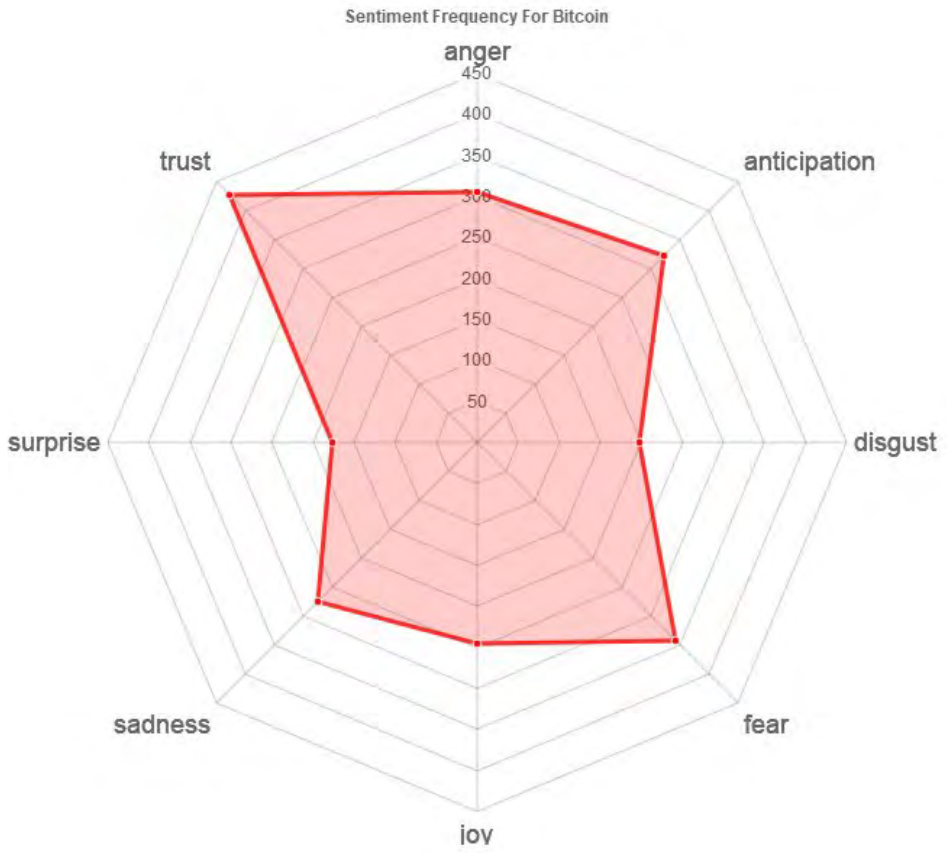
Bitcoin - NRC Sentiment Bar Chart

```
btc_sentiment2 <- tweet_final %>% inner_join(get_sentiments("nrc")) %>% count(word, sentiment)
topwords <- btc_sentiment2 %>%  group_by(sentiment) %>%  top_n(10) %>% ungroup() %>%  mutate(word = reorder(word, n))
ggplot(topwords, aes(word, n, fill= sentiment)) +
        geom_col(show.legend = FALSE) + facet_wrap(~sentiment, scales = "free") +
        labs(title= "Bitcoin Word Trend Tweets") + coord_flip()
```



NRC Radar Chart

```
nrc_radar_chart <- btc_sentiment2 %>% filter(!grepl("positive|negative", sentiment))
word_tally <- nrc_radar_chart %>% group_by(sentiment) %>% tally()
chartJSRadar(scores= word_tally, showLegend = FALSE, main= "Sentiment Frequency For Bitcoin")
```



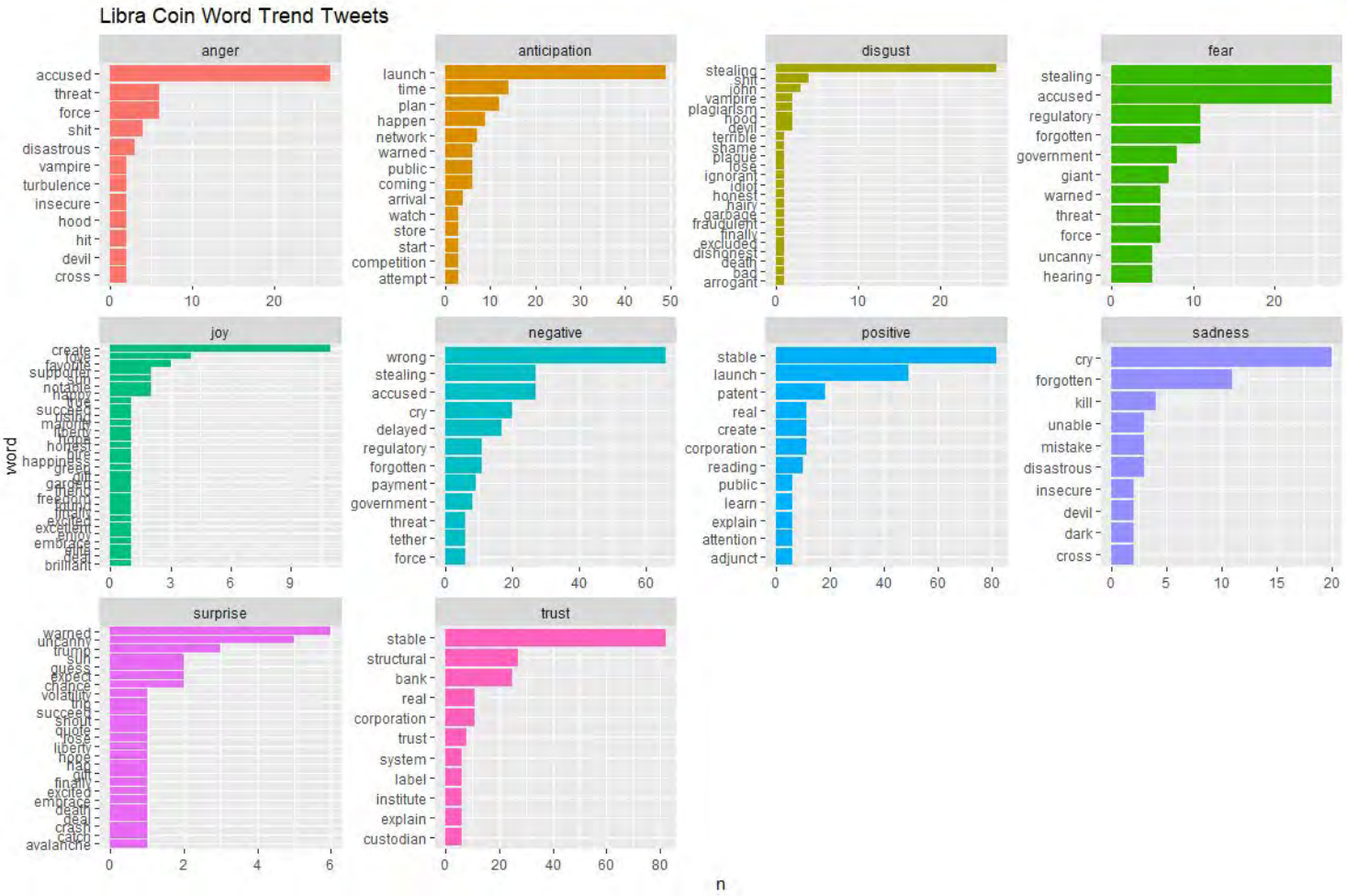
From the radar graph of “Bitcoin” we can be relatively confident that the results are consistent with those of “Cryptocurrency”

What about Libra Coin?

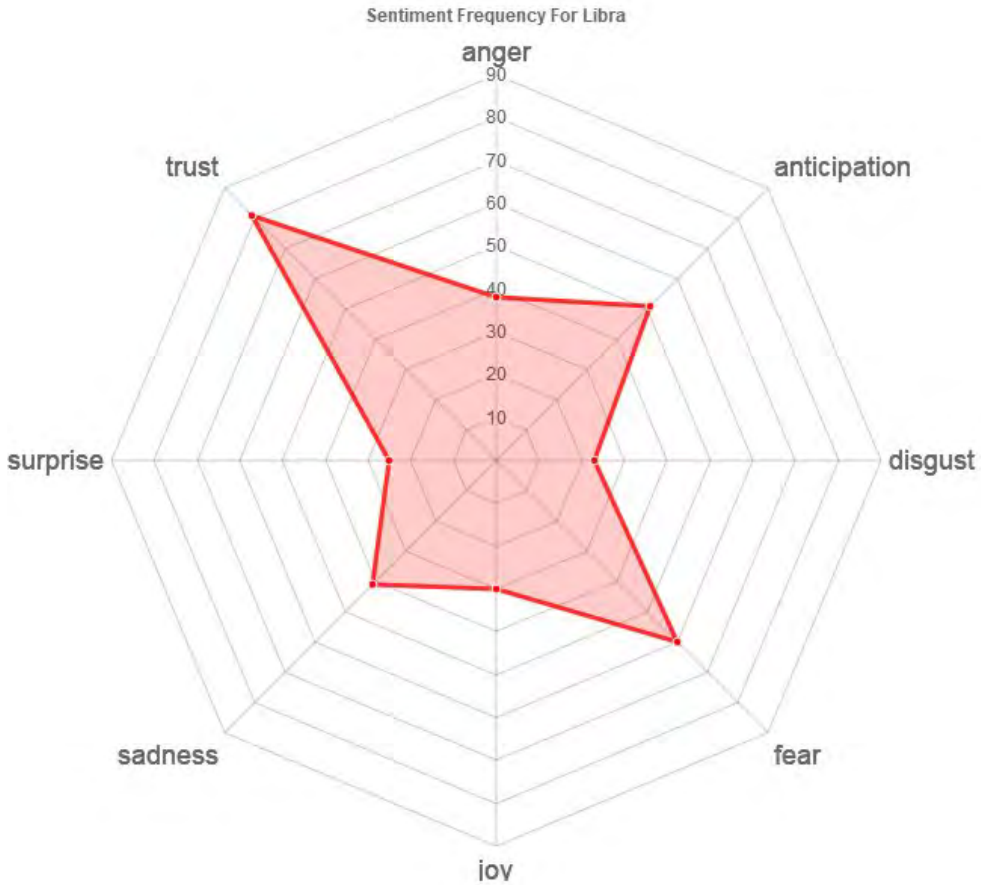
```
libra_twitter <- searchTwitter("libra coin",n=1196,lang="en")
libra_twitter_df <- twListToDF(libra_twitter)
tweet_words <- libra_twitter_df %>% select(id, text) %>% unnest_tokens(word,text)
my_stop_words <- stop_words %>% select(-lexicon) %>%
        bind_rows(data.frame(word = c("a","and","in","is","https","for","of","on","partridge","rt","t.co",
                                      "this","that","trent","trentpartridge","with","y","you")))
tweet_words_interesting <- tweet_words %>% anti_join(my_stop_words)

my_stop_words <- tibble(word = c("1","2","3","4","advent","amp","angel","cap","cash","congress","credit","exchange","follow","giveaway",
                "money","music","president","pretty","professor","save","share","technology","trade"),lexicon ="twitter" )
new_stop_list <- stop_words %>% bind_rows(my_stop_words)
tweet_final <- tweet_words_interesting %>% anti_join(new_stop_list)
```

Libra - NRC Sentiment Bar Chart

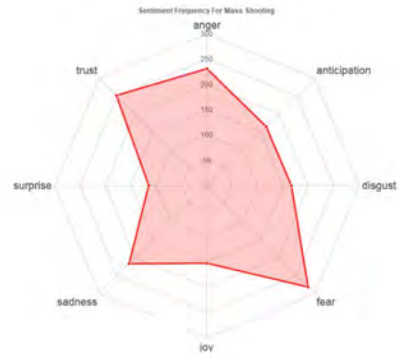
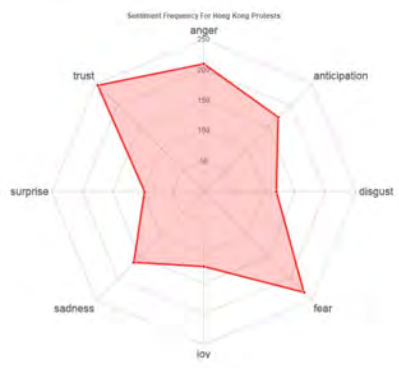
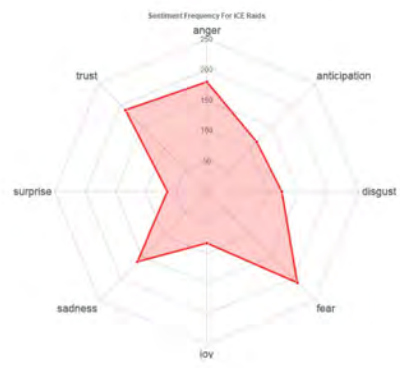


Libra - NRC Radar Chart

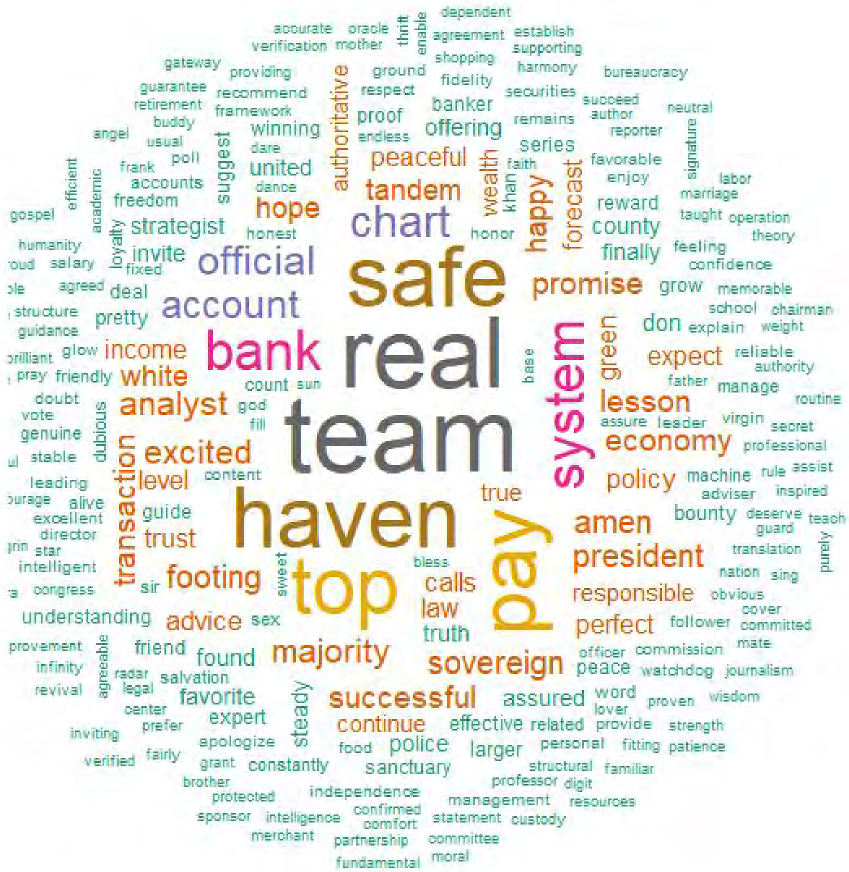


Before any suggestion, I want to point out that due to technical issues, Twitter only allow me to fetch 1196 tweets about “Libra Coin”, and I’m still not sure if some of them include information about astrology. Nonetheless, results from the radar chart stay pretty consistent with the previous two. However, the contradiction between trust and fear is still to be further examined. The anger and joy parameter also shrink in this case, indicating higher percentage in other sections.

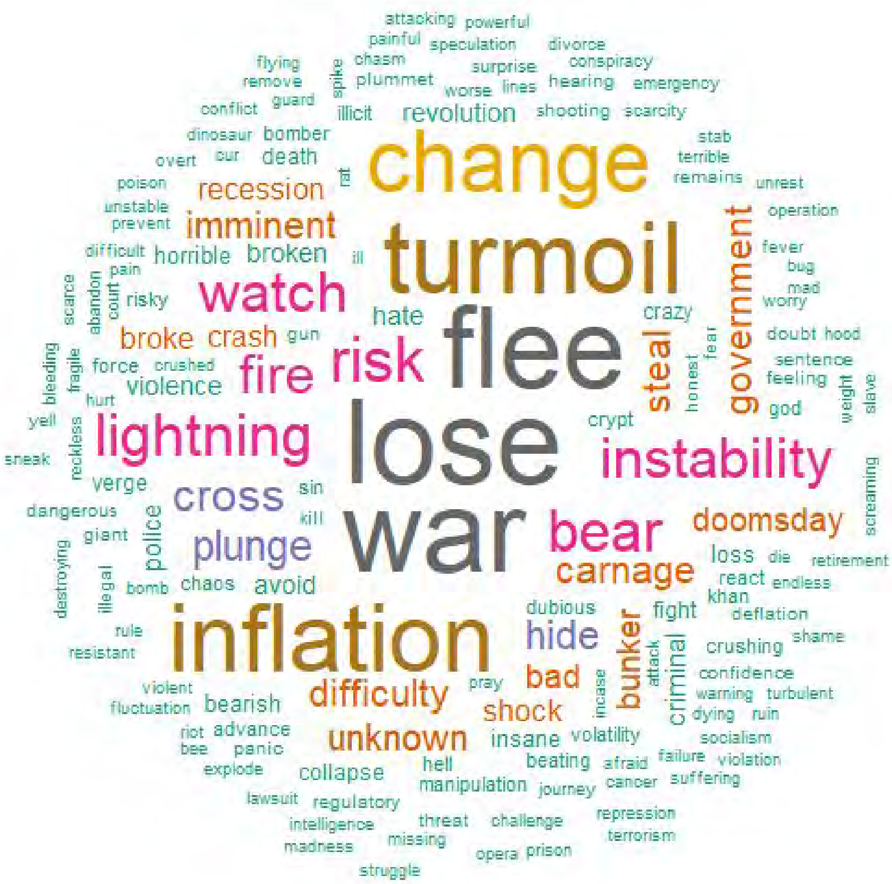
For comparison, These are the sentiment radar chart for recent incidents

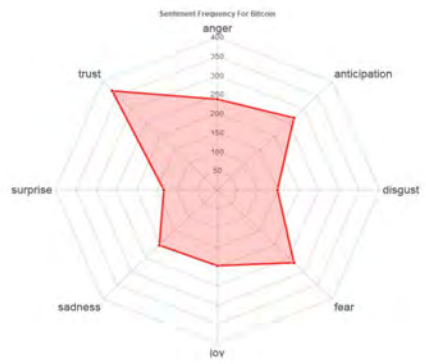
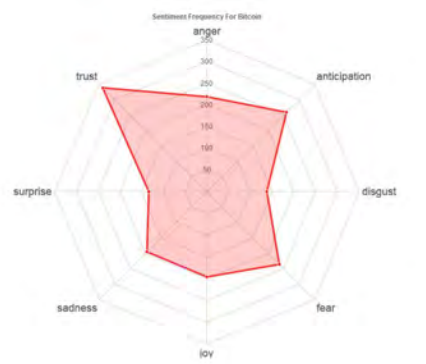
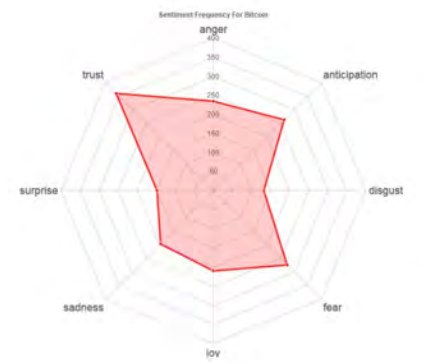
Trade War	**Mass Shooting (shortly after the El Paso, Dayton and Chicago incidents)**
	
Hong Kong Protests	**ICE Raids**
	

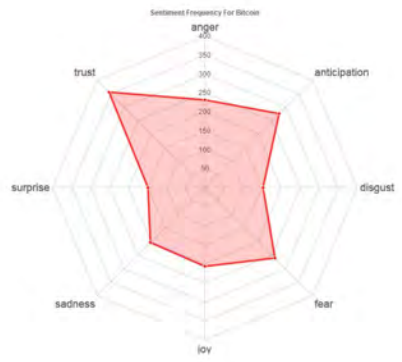
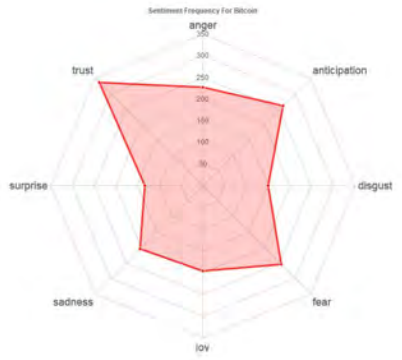
Why people trust Bitcoin?

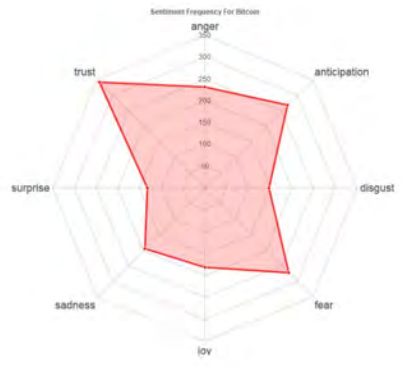
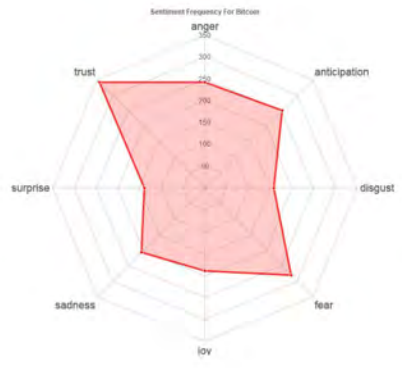


Why people fear Bitcoin?



2019-07-30	Notes
	● BTC-USD up 5.15%
2019-07-31	
	● Less anger and higher trust ● BTC-USD up 3.18%
2019-08-01	
	● Higher anticipation ● BTC-USD up 1.15%
2019-08-02	
	● Less sadness and higher trust ● BTC-USD up 2.77%

2019-08-03	Notes
	● BTC-USD up 1.46%
2019-08-04	
	● BTC-USD up 7.55%
2019-08-05	
	● BTC-USD down 2.89%
2019-08-06	
	● More anger and more trust ● BTC-USD up 4.42%

2019-08-07	Notes
	• More trust and more fear • BTC-USD up 0.07%
2019-08-08	
	• BTC-USD down 0.75%