

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

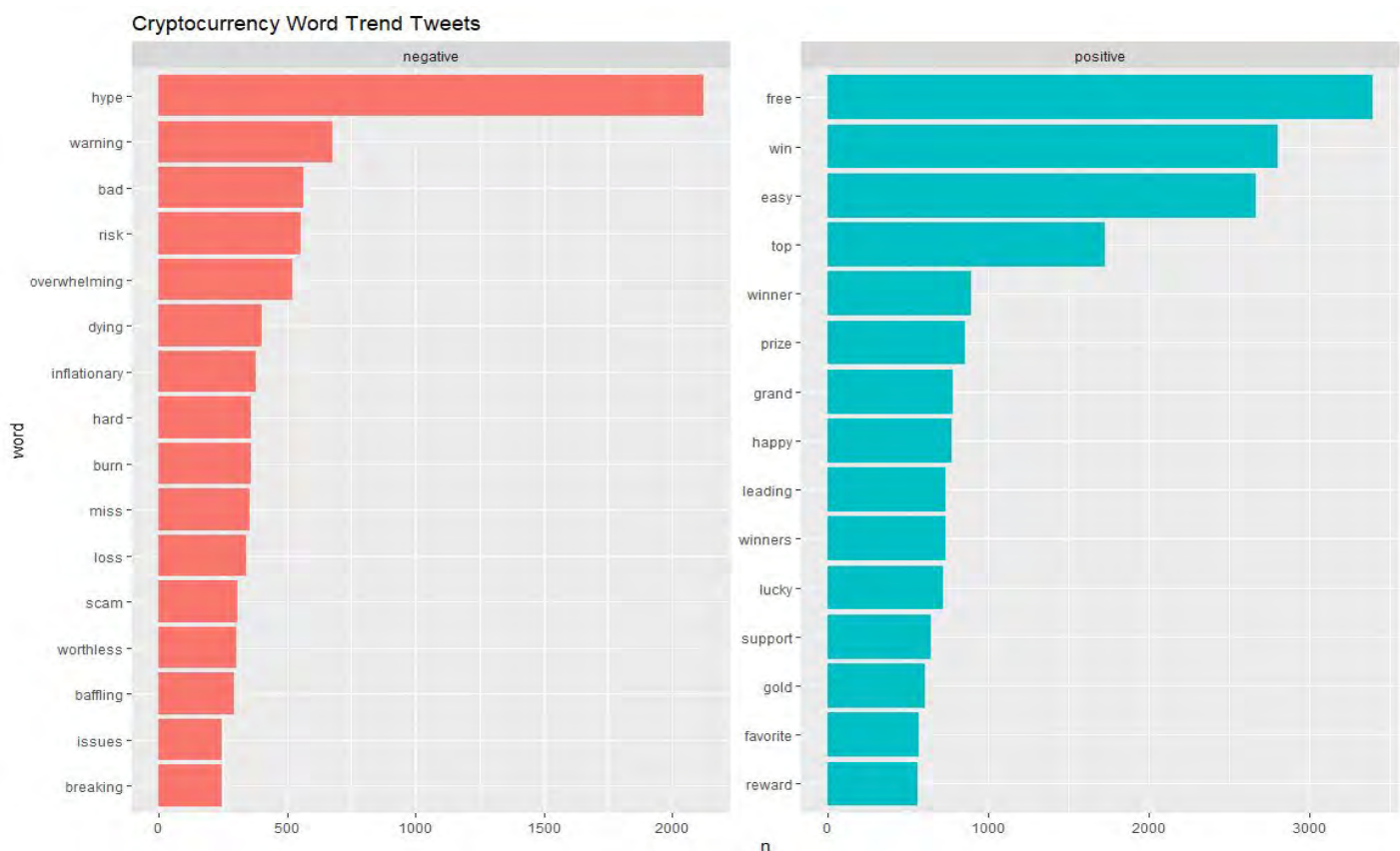
```
my_stop_words <- stop_words %>% select(-lexicon) %>%
  bind_rows(data.frame(word = c("a","and","in","is","https","for","of","on","partridge","rt","t.co",
    "this","that","trent","trentpartridge","with","y","you")))
tweet_words_interesting <- tweet_words %>% anti_join(my_stop_words)
tweet_words_interesting %>% 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 stopwords 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
```

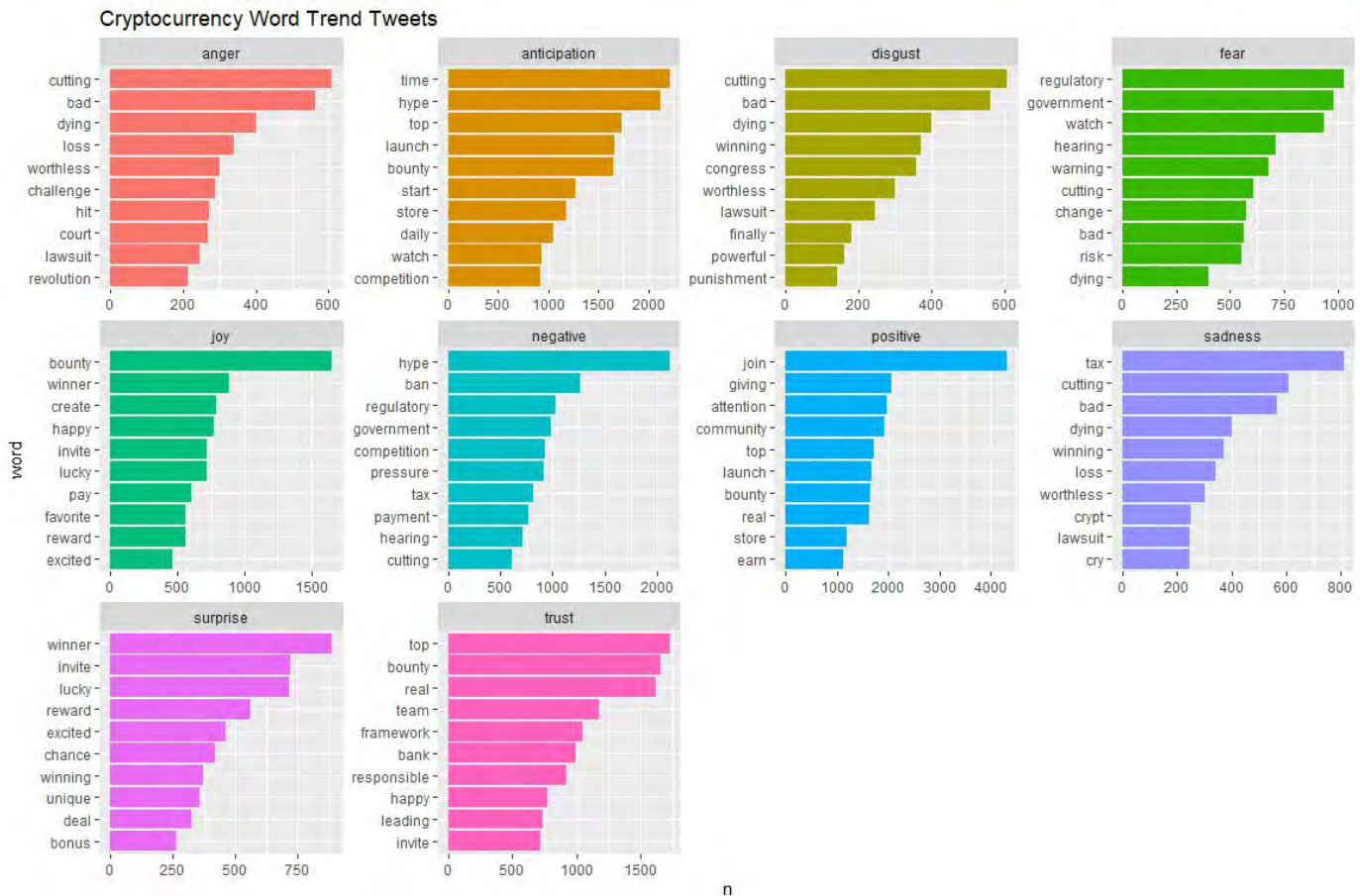
Bing Sentiment Analysis

```
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()
```



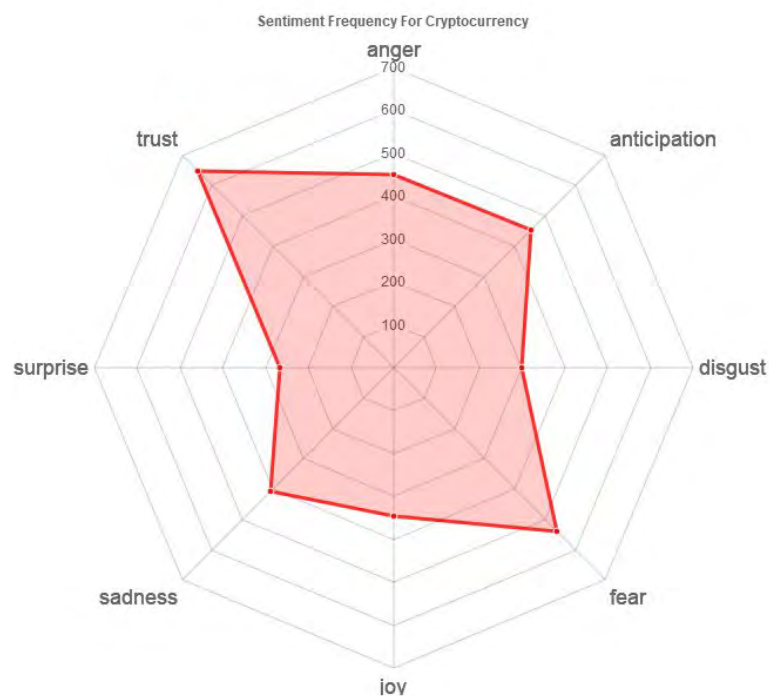
Cryptocurrency - NRC Sentiment Analysis

```
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.

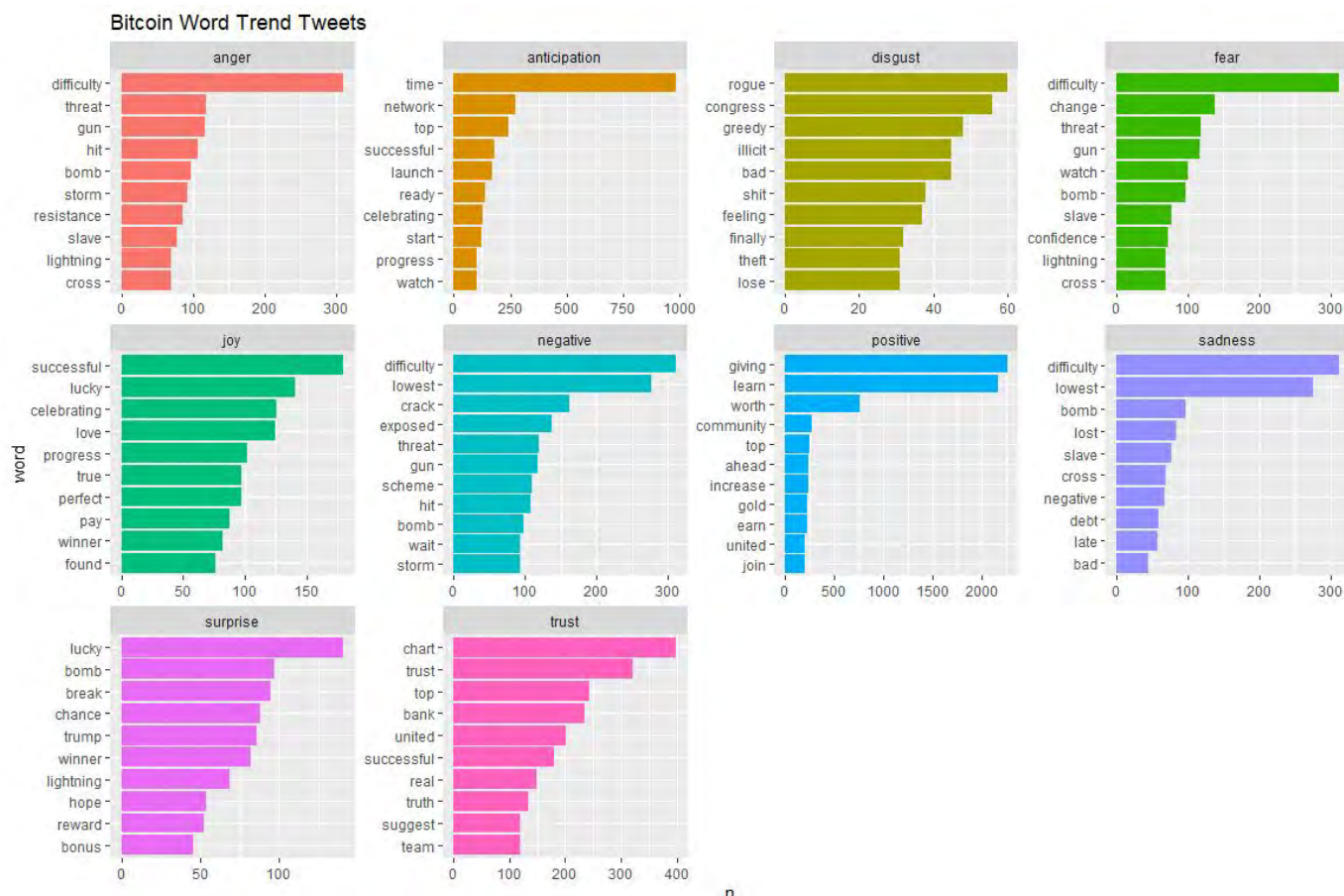
What about Bitcoin?

```
btc_twitter <- searchTwitter("bitcoin",n=20000,lang="en")
btc_twitter_df <- twListToDF(btc_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)
```

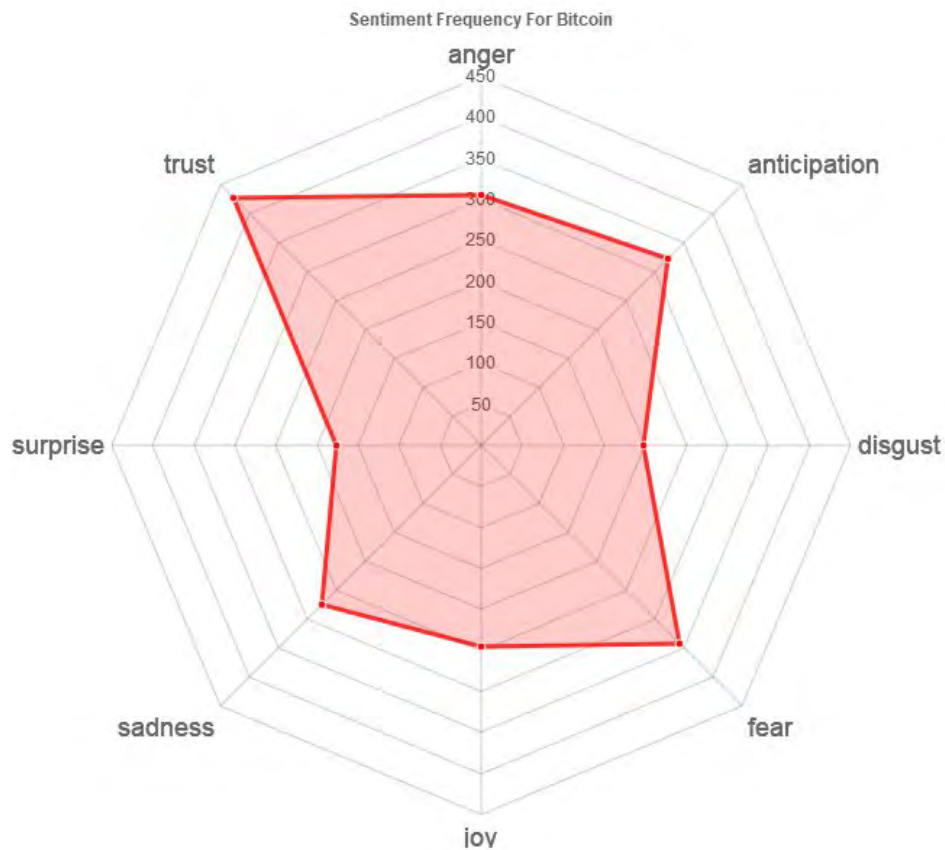
Bitcoin - NRC Sentiment Bar Chart

```
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")
```



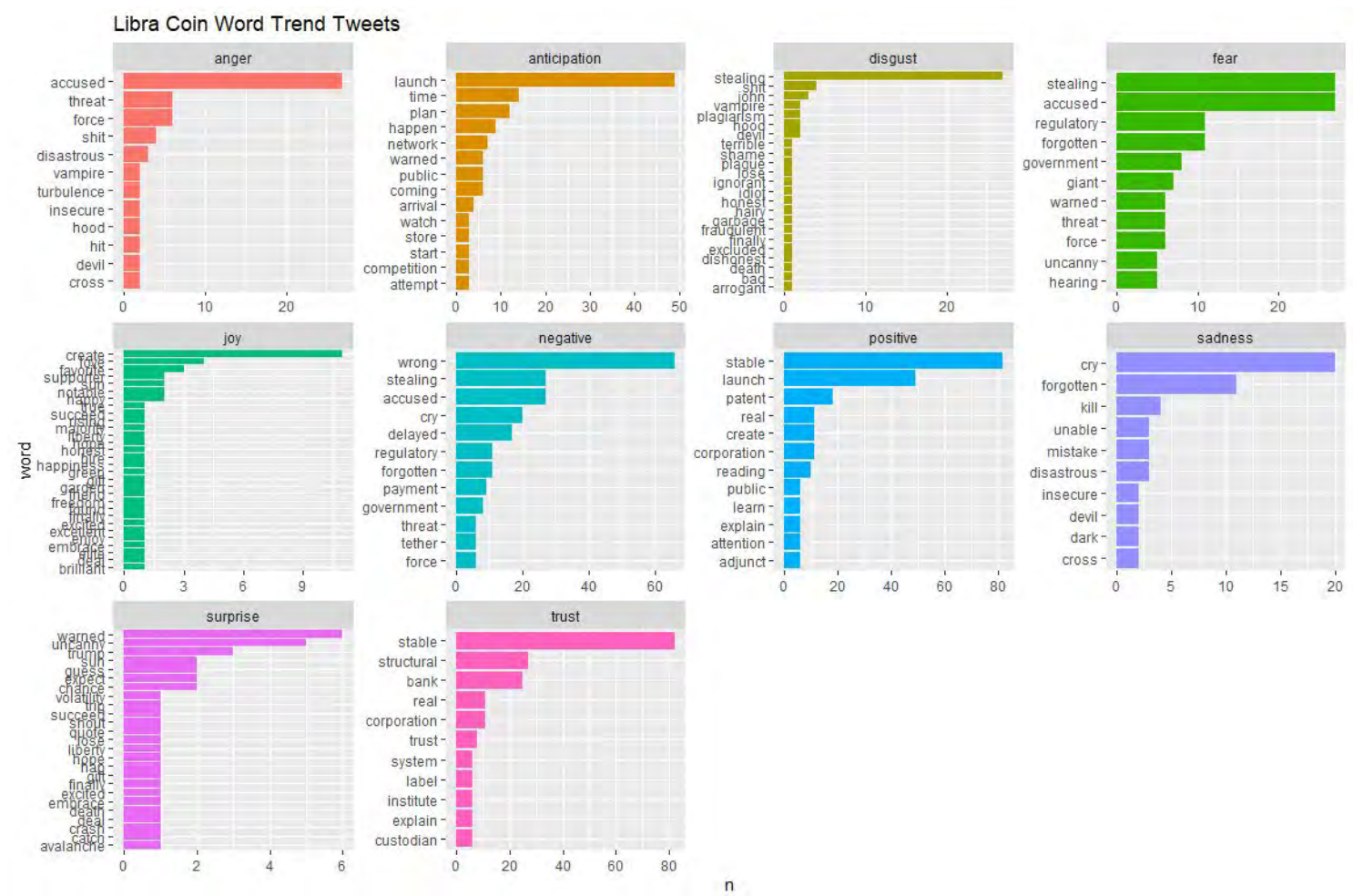
From the radar graph of “Bitcoin” we can be relatively confident that the results are consistent with those of “Cryptocurrency”

What about Libra Coin?

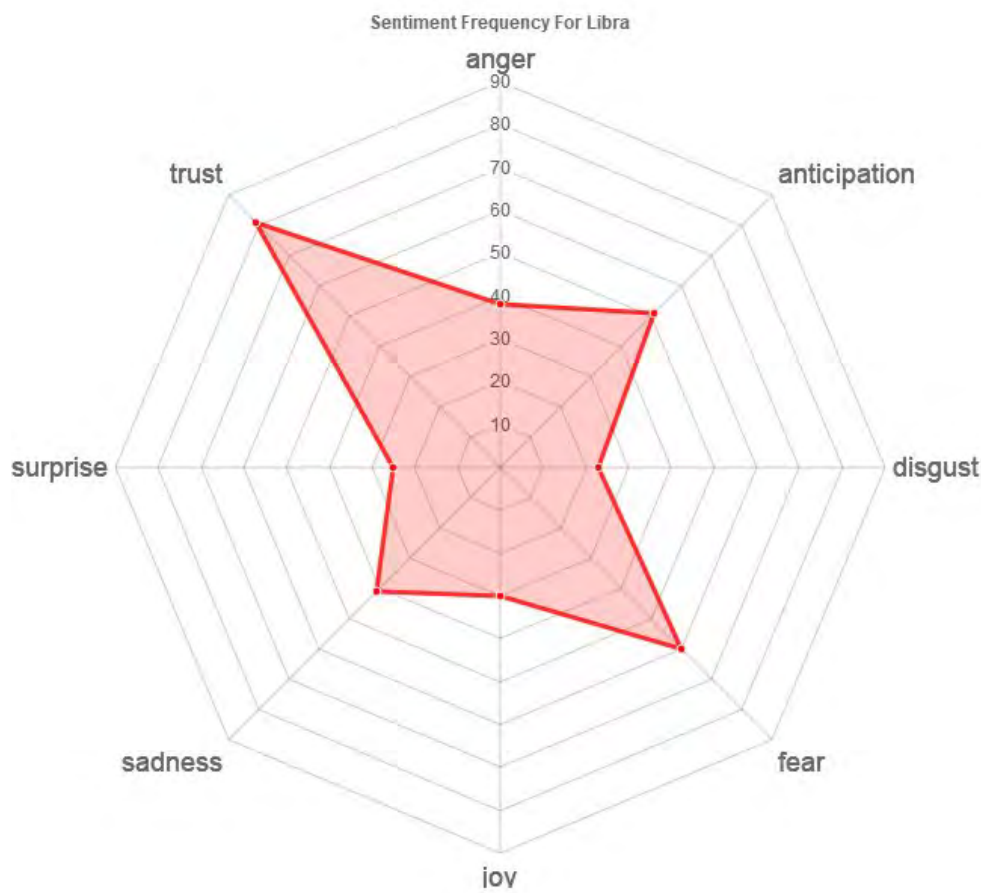
```
libra_twitter <- searchTwitter("libra coin",n=11196,lang="en")
libra_twitter_df <- twListToDF(libra_twitter)
tweet_words <- libra_twitter_df %>% select(id, text) %>% unnest_tokens(word,text)
my_stop_words <- stop_words %>% select(-lexicon) %>%
  bind_rows(data.frame(word = c("a","and","in","is","https","for","of","on","partridge","rt","t.co",
    "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


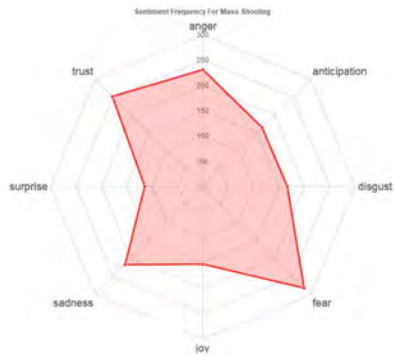
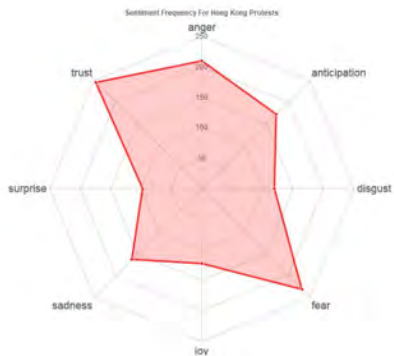
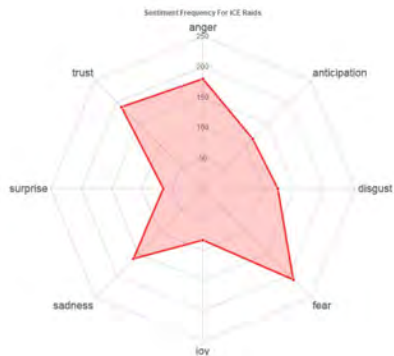


Libra - NRC Radar Chart

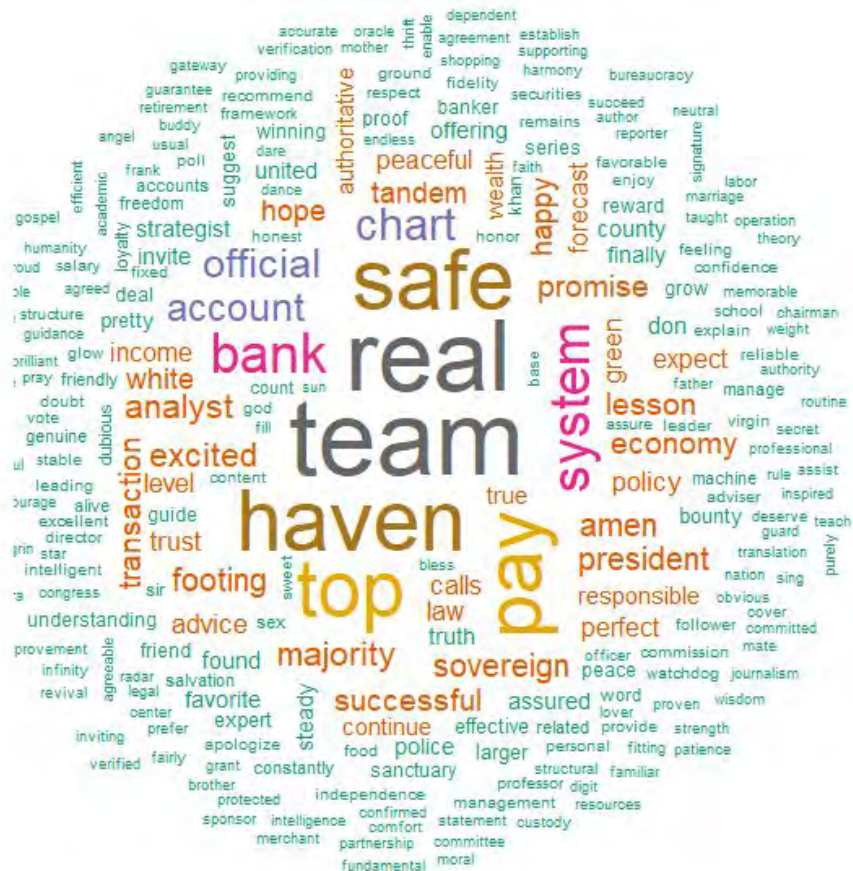


Before any suggestion, I want to point out that due to technical issues, Twitter only allow me to fetch 1196 tweets about “Libra Coin”, and I’m still not sure if some of them include information about astrology. Nonetheless, results from the radar chart stay pretty consistent with the previous two. However, 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.

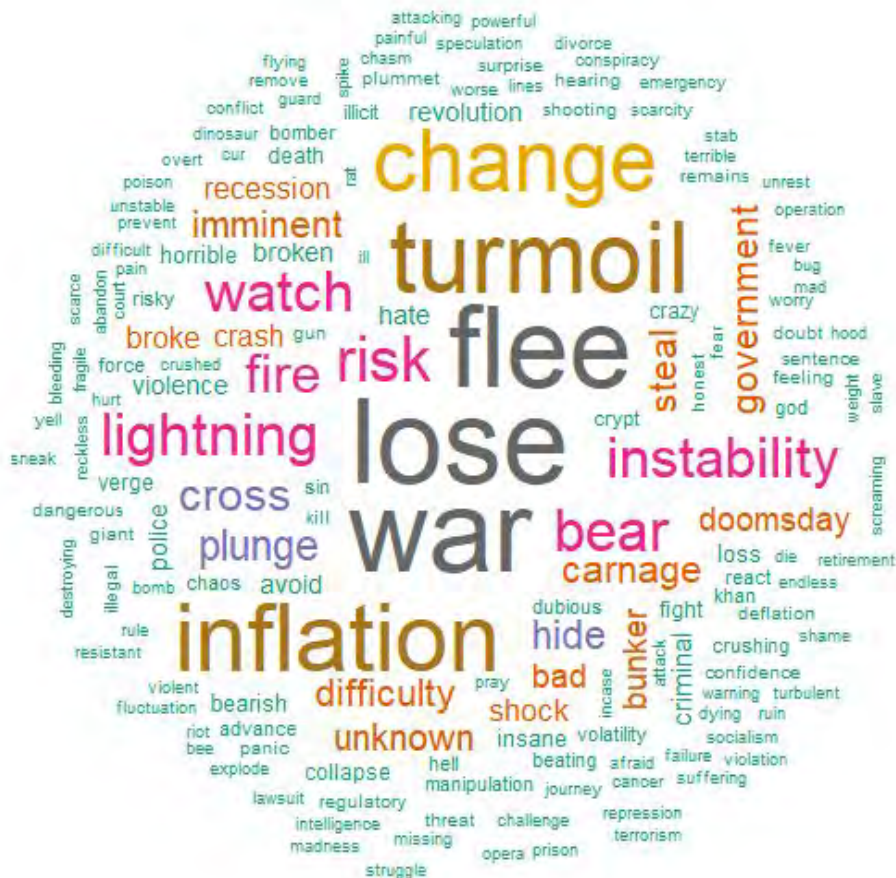
For comparison, These are the sentiment radar chart for recent incidents

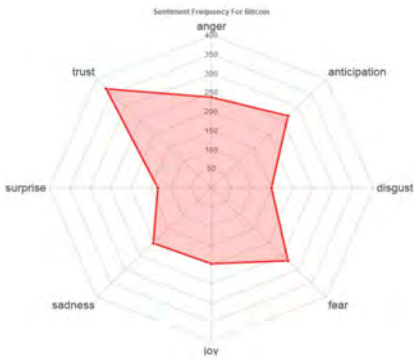

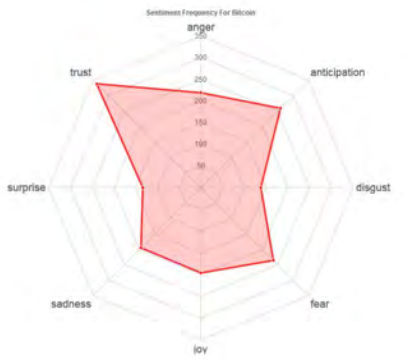
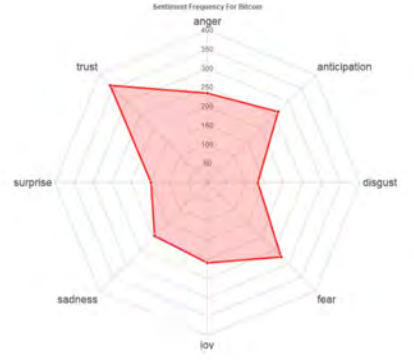
Trade War	Mass Shooting (shortly after the El Paso, Dayton and Chicago incidents)
 <p>Sentiment Frequency for Trade War</p> <p>This radar chart displays sentiment frequencies for the Trade War across eight categories: anger, anticipation, disgust, fear, joy, sadness, surprise, and trust. The chart shows a prominent peak in 'fear' (approx. 220) and 'disgust' (approx. 180). 'Anger' is also significant at approx. 150. 'Trust' and 'joy' are the lowest, both around 50. 'Surprise' and 'sadness' are also low, around 70-80.</p>	 <p>Sentiment Frequency for Mass Shooting</p> <p>This radar chart displays sentiment frequencies for mass shootings across eight categories. 'Fear' is the dominant sentiment at approx. 250. 'Disgust' follows at approx. 180. 'Anger' is at approx. 150. 'Trust' is around 100, while 'joy', 'surprise', and 'sadness' are the lowest, all below 50.</p>
Hong Kong Protests	ICE Raids
 <p>Sentiment Frequency for Hong Kong Protests</p> <p>This radar chart displays sentiment frequencies for Hong Kong protests across eight categories. 'Fear' is the highest at approx. 220. 'Disgust' is at approx. 180. 'Anger' is around 150. 'Trust' is at approx. 100. 'Joy', 'surprise', and 'sadness' are the lowest, all around 50.</p>	 <p>Sentiment Frequency for ICE Raids</p> <p>This radar chart displays sentiment frequencies for ICE raids across eight categories. 'Fear' is the highest at approx. 200. 'Disgust' is at approx. 150. 'Anger' is around 100. 'Trust' is at approx. 100. 'Joy', 'surprise', and 'sadness' are the lowest, all around 50.</p>


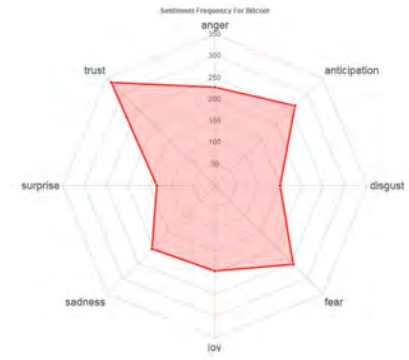
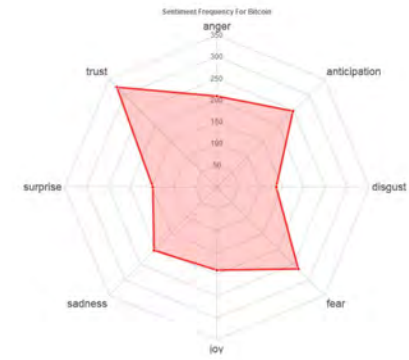
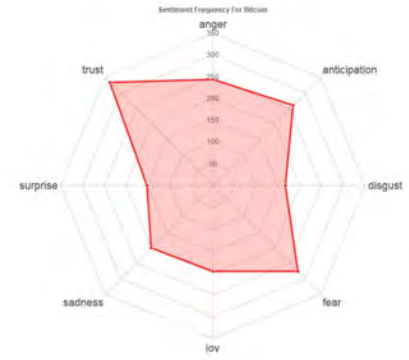
Why people trust Bitcoin?

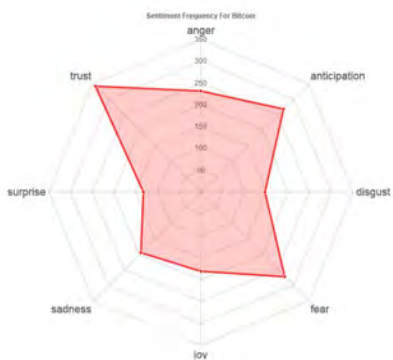
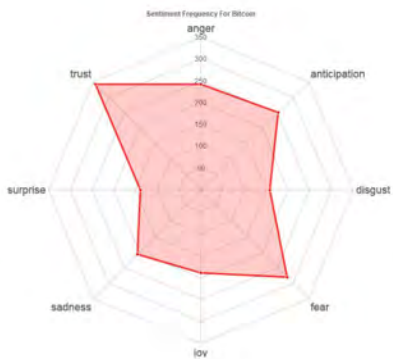


Why people fear Bitcoin?



2019-07-30	Notes
 <p>A radar chart titled 'Sentiment Frequency for Bitcoin' showing sentiment levels across eight categories: anger, anticipation, disgust, fear, joy, sadness, surprise, and trust. The chart has concentric grid lines at intervals of 50, ranging from 0 to 400. The data is represented by a red-filled area. Trust is the highest sentiment at approximately 250, followed by anticipation at 200. Disgust and fear are both at 100, while anger, surprise, and joy are all at 50.</p>	<ul style="list-style-type: none"> ● BTC-USD up 5.15%
2019-07-31	
 <p>A radar chart titled 'Sentiment Frequency for Bitcoin' showing sentiment levels across eight categories: anger, anticipation, disgust, fear, joy, sadness, surprise, and trust. The chart has concentric grid lines at intervals of 50, ranging from 0 to 400. The data is represented by a red-filled area. Trust is the highest sentiment at approximately 250, followed by anticipation at 200. Disgust and fear are both at 100, while anger, surprise, and joy are all at 50.</p>	<ul style="list-style-type: none"> ● Less anger and higher trust ● BTC-USD up 3.18%
2019-08-01	
 <p>A radar chart titled 'Sentiment Frequency for Bitcoin' showing sentiment levels across eight categories: anger, anticipation, disgust, fear, joy, sadness, surprise, and trust. The chart has concentric grid lines at intervals of 50, ranging from 0 to 400. The data is represented by a red-filled area. Trust is the highest sentiment at approximately 250, followed by anticipation at 200. Disgust and fear are both at 100, while anger, surprise, and joy are all at 50.</p>	<ul style="list-style-type: none"> ● Higher anticipation ● BTC-USD up 1.15%
2019-08-02	
 <p>A radar chart titled 'Sentiment Frequency for Bitcoin' showing sentiment levels across eight categories: anger, anticipation, disgust, fear, joy, sadness, surprise, and trust. The chart has concentric grid lines at intervals of 50, ranging from 0 to 400. The data is represented by a red-filled area. Trust is the highest sentiment at approximately 250, followed by anticipation at 200. Disgust and fear are both at 100, while anger, surprise, and joy are all at 50.</p>	<ul style="list-style-type: none"> ● Less sadness and higher trust ● BTC-USD up 2.77%

2019-08-03	Notes
	<ul style="list-style-type: none"> BTC-USD up 1.46%
2019-08-04	
	<ul style="list-style-type: none"> BTC-USD up 7.55%
2019-08-05	
	<ul style="list-style-type: none"> BTC-USD down 2.89%
2019-08-06	
	<ul style="list-style-type: none"> More anger and more trust BTC-USD up 4.42%

2019-08-07	Notes																		
 <p>A radar chart titled "Sentiment Frequency for Bitcoin" showing sentiment levels across eight categories: anger, anticipation, disgust, fear, joy, sadness, surprise, and trust. The chart has concentric rings representing frequency from 0 to 350. The data series is a red-filled polygon. Trust is the highest at approximately 280, followed by anticipation at 220, fear at 200, and joy at 180. Anger, surprise, and sadness are the lowest, all around 100.</p> <table border="1"><thead><tr><th>Sentiment</th><th>Frequency (approx.)</th></tr></thead><tbody><tr><td>anger</td><td>100</td></tr><tr><td>anticipation</td><td>220</td></tr><tr><td>disgust</td><td>100</td></tr><tr><td>fear</td><td>200</td></tr><tr><td>joy</td><td>180</td></tr><tr><td>sadness</td><td>100</td></tr><tr><td>surprise</td><td>100</td></tr><tr><td>trust</td><td>280</td></tr></tbody></table>	Sentiment	Frequency (approx.)	anger	100	anticipation	220	disgust	100	fear	200	joy	180	sadness	100	surprise	100	trust	280	<ul style="list-style-type: none">• More trust and more fear• BTC-USD up 0.07%
Sentiment	Frequency (approx.)																		
anger	100																		
anticipation	220																		
disgust	100																		
fear	200																		
joy	180																		
sadness	100																		
surprise	100																		
trust	280																		
2019-08-08																			
 <p>A radar chart titled "Sentiment Frequency for Bitcoin" showing sentiment levels across eight categories: anger, anticipation, disgust, fear, joy, sadness, surprise, and trust. The chart has concentric rings representing frequency from 0 to 350. The data series is a red-filled polygon. Trust is the highest at approximately 280, followed by anticipation at 220, fear at 200, and joy at 180. Anger, surprise, and sadness are the lowest, all around 100.</p> <table border="1"><thead><tr><th>Sentiment</th><th>Frequency (approx.)</th></tr></thead><tbody><tr><td>anger</td><td>100</td></tr><tr><td>anticipation</td><td>220</td></tr><tr><td>disgust</td><td>100</td></tr><tr><td>fear</td><td>200</td></tr><tr><td>joy</td><td>180</td></tr><tr><td>sadness</td><td>100</td></tr><tr><td>surprise</td><td>100</td></tr><tr><td>trust</td><td>280</td></tr></tbody></table>	Sentiment	Frequency (approx.)	anger	100	anticipation	220	disgust	100	fear	200	joy	180	sadness	100	surprise	100	trust	280	<ul style="list-style-type: none">• BTC-USD down 0.75%
Sentiment	Frequency (approx.)																		
anger	100																		
anticipation	220																		
disgust	100																		
fear	200																		
joy	180																		
sadness	100																		
surprise	100																		
trust	280																		