

Problem Set: ARM

You will continue to use the TWITTER API to collect tweets and analyze them. This week you will find associations between words found in sets of tweets, using ARM.

1. Run the ARM coded example provided and choose a hashtag of your liking and collect 100 tweets. What hashtag did you choose?
2. Create a wordcloud of all words from all tweets collected, and identify words that are non-informative and should be removed for the purposes of analysis (stopwords). What stopwords did you choose.
3. Create transaction data from the tweets and identify 5 rules that are “most interesting”. What criteria did you use to determine whether a rule was “interesting”? Hint: Lift, Support, Confidence, ... ?
4. Create a visualization (using arulesviz or another) of the rules found as a result of your ARM model.