1. **Initial Questions/Why are we doing this?**

What are they saying? Why are they saying it?

-Are they mentioning a FI, not mentioning a FI, comparing multiple FIs?

-Are they mentioning a specific product or service?

-Both of the above would use Keyword searches or identify who the user is tweeting @

-Was it a positive or negative experience (can possibly scale based on degree)

-Sentiment Analysis

-Why was it a positive/negative experience (e.g. issues with tech, customer service, change or discontinuing of a product, price sensitivity, etc.)

Who – Tweeters

-Demographics (Wealth, Age, gender, race, single/married, etc.

-Types of Consumers (Builders, Savers, Planners, Balance sheet managers)

-Both would need to be classified through body of work of their other tweets or network characteristics (friends/followers/retweets) **(Authorship Analysis)**

-We could use authorship analysis that has been done (academic papers, etc.) or do our own, but that would require a two-step process (Classifying linguistics based on information that we already know, and then applying it to the desired body of tweets)

-Or find this through an automated sampling process to complainer/complimenters

-Who / How Many people are viewing the tweets? (See network effects)

Where

-Region

-Country

-Possibly down to branch

-Developed/Emerging

-?Can all be found via zip code geolocater (?)

When

-Are trends at all cyclical (people comment based on time of day, week, month, or pay period)?

-Do people comment differently around seasons/holidays/time of year?

-Do people comment differently around events pertaining to financial institution (new product, change in service, change in infrastructure, etc.)? Are they reacting positively or negatively to change?

-Do people comment differently around events pertaining to their own lives (get married, have kids, graduate school, new job, new mortgage, etc.)?

-This would require looking at past tweets or survey sampling

------------------------------------------------------------------------------------------------------------------------------------------

**2. Analytical Methods**

-Frequency Analysis:

-Used primarily in What and Why questions

-What issues are commonly cited by users? What institutions are commonly being mentioned? What issues are associated with each institution? How often are services and institutions being compared to one another?

-Comparison wordcloud: words specific to one FI/topic on the edge, words common to multiple ones gravitate towards center

-Wordcloud talking about some given topic:

-creates groups based on frequency of words used together (?)

-Ex: Given topic: Bank of America

-Groups: checking, loans, mortgages, online customer experience

-Terms: bad, difficult, expensive, understanding

-Modified tag-cloud: hot-and-cold given two tweeters/topics

-Mining entities from user’s tweets (some of this relates to Authorship Analysis)

-Average # words per tweet

-Average word length

-Number of # per tweet

-Lexical diversity

**-Most frequent words/terms?**

-What hashtags they use

-Who they tweet at

-Who retweets them/ who do they retweet

------------------------------------------------------------------------------------------------------------------------------------------

-Sentiment Analysis

-Used primarily in What and Why questions

-Do users feel positively or negatively about certain institutions and products/services? How strongly do they feel about it? What emotions (anger/confusion/surprise) are associated with their experiences?

-Opinion = (oj, fjk, soijkl, hi, tl)

- oj = the product

- fjk = feature of the product

- soijkl = sentiment value of the opinion of the opinion holder hi on feature fjk of object oj at time tl

-ex: (Bank of America, online checking, difficult (=-1), 31-year old white married mass-affluent male who classifies as a saver, right after a tweak to the service)

-Part-of-speech tagging (plus position and more) labels each word

*-This however is based on formal language. On twitter, there are so many misspellings, abbreviations, and sentence fragments that I think this would be difficult*

-Machine tries to situate words on an emotive scale (Far negative to far positive)

-more degrees means less accuracy

-Breen’s approach: Develop a score for each tweet. Eg. Score = number of positive words – number of negative words

-Need to import documents with list of positive and negative words

-Harvest tweets under certain criteria (by mention, user, hashtag, etc.)

-R sentiment package: algorithms describe text either by polarity (positive/negative) or by emotion

-*Separating by emotion could be of value in answering the “Why” question. E.g.* *Are they angry, annoyed, confused, etc. with a particular product or aspect of service?*

-Viralheat’s Sentiment Analysis API: uses algorithms to give a positive/negative sentiment and score of -1 to 1.

-Examples were remarkably inaccurate

*Both of these seem very doable. Examples (*[*https://sites.google.com/site/miningtwitter/questions/sentiment*](https://sites.google.com/site/miningtwitter/questions/sentiment)*) are understandable, and can replace drinks/Starbucks with members.*

*I see the biggest issue being positive/negative words or emotion/polarity attributed to something else in the tweet and not to the product*

*For example, in the second method the tweet “chad ocho cinco mad as hell some one took his starbucks card …” registers an emotion of ‘anger’ and a polarity ‘negative’. Most likely, the algorithm picked up words like “mad” and “hell” and registered negative sentiment. Using Breen’s approach, “mad” and “hell” would again likely fall under the negative words document, and would pick up a score of -2. However, while qualitatively looking at the tweet, the fact that Chad is upset that someone took his card means the card meant something to him, which would be good for Starbucks’ brand image.*

*Additionally, many tweets register as unknown emotion, but the algorithm does better on what is considered positive, negative, or neutral.*

*Keep in mind bias vs. inaccuracy*

------------------------------------------------------------------------------------------------------------------------------------------

Authorship Analysis:

-Used for Who questions

-What demographic groups does each user fall under? What type of consumer are they?

<http://www.academia.edu/6422757/Gender_Prediction_on_Twitter_Using_Stream_Algorithms_with_N-Gram_Character_Features>

-Demographic indicators based on speech patterns (Authorship analysis)

-Particular study: Identify gender using Perceptrion and Naïve Bayes with 1-5 gram features from tweet text

-Gram: continuous string of characters, words, syllables, etc. Advantage over a dictionary due to misspellings, acronyms, emoticons

-Naïve Bayes algorithm calculates a probability from the occurrence of each feature, with regards to the true class of the instance.

-Misspellings, acronyms, and emoticons can be used to identify author characteristics

-All of this is a two-step approach: Use data where you know the author characteristics to match text features to demographic groups, and then use those features to predict tweets with unknown author characteristics

<http://www.cs.jhu.edu/~delip/smuc.pdf>

Network structure: See if different groups have different distribution of # of followers, # following, and follower-following ratio

SocioLinguistic-feature Models: word choice, emoticons, abbreviations attributable to different demographic groups

Possesives: using “my” in front of a word distinguishes author characteristics

-ex: Male: “my wife”, “my shorts”. Female: “My husband”, “my dress”

------------------------------------------------------------------------------------------------------------------------------------------

Network Effects:

-Used to answer Who questions

-Who is the user connected to? Who is seeing the author’s tweet? 🡪 What impact does the tweet have?

-Use lists of friends, followers/following, retweets, and favorites to build a network for each user

-Size and importance of user’s network projects how many people see a tweet and how much impact it generates

-Similar to Klout score

Survey Sampling:

-Automate a process where users are sent a sampling survey to complete after a relevant tweet. Could be sent by CEB, but may be more effective if sent by members who are being mentioned or tweeted @ (see Unintended Benefit below)

-E.g. Process can be sent to users who mention or tweet @ a financial institution

-Can start with ‘What’ and ‘Why’ questions. If users were willing to voice opinion in a public forum, hopefully they’d be willing to give feedback to improve their experience

-Would reduce the need of keyword searches and sentiment analysis

-Unintended benefit: users will be happy that financial institution is asking for their feedback and trying to improve its processes and services

-Can include ‘Who’ questions about demographics and the type of consumer they are to limit the need for authorship analysis

-Since the typical user’s Twitter experience is quick and ad hoc use, ideally, the survey would be sent ASAP and would get to the point quickly

**3. Building a Base of Tweets:**

-Step 1: Tweets mentioning members or @members

-After that, 2 approaches: Broad vs. Narrow

-Broad: Large body of tweets, a lot of noise

-Narrow: Smaller body of tweets, hopefully more focused on our objective

-Building a Lexicon:

Manually

*-Term List:* Bank, atm, mortgage, loan, credit card, finance, insurance, invest, lend, owe, profit, refund

-Contributions from program teams

Find Existing One

Huge lexicons on lexicon.ft.com, Investopedia

-Too big, not commonly used terms

-On the other hand, people probably aren’t tweeting about “consumer credit files”, so will not likely cause to much noise

-http://www.wjh.harvard.edu/~inquirer/Econ%40.html

-http://www.wjh.harvard.edu/~inquirer/Exch.html (smaller list than above)

-These have modifiers that seem to affect the context of the use

-Track down methodologies that others have used

-Build Out:

**Step 6:** What are the most frequent words

# most frequent words

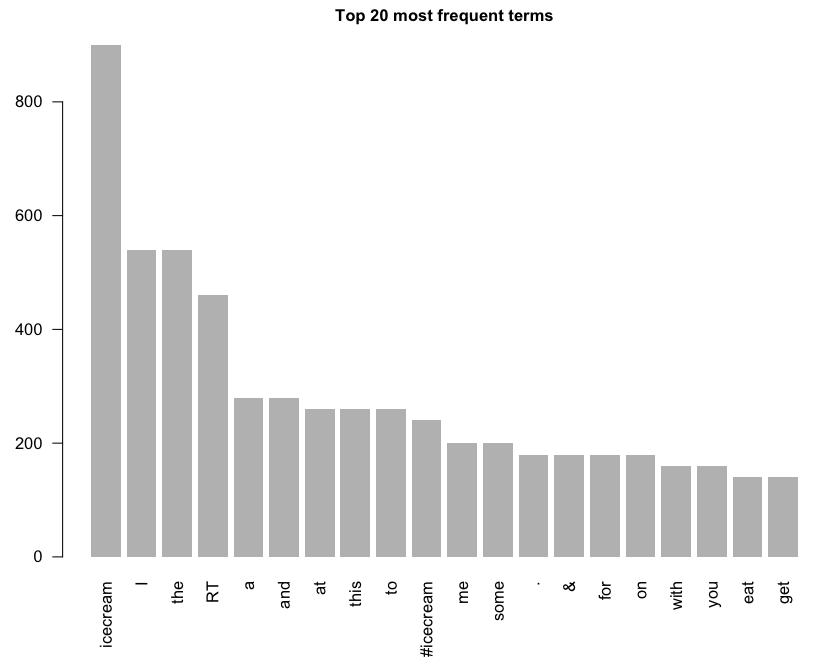
mfw = sort(table(unlist(words\_list)), decreasing=TRUE)

# top-20 most frequent

top20 = head(mfw, 20)

# barplot

barplot(top20, border=NA, las=2, main="Top 20 most frequent terms", cex.main=1)

Problem: get a result like this: 

-Or-

**-Step 4:** Create Lexical Corpus and term-document matrix

# create corpus

corpus = Corpus(VectorSource(results))

# remove stopwords

skipwords = c(stopwords("english"),

   "genetics", "genomics", "genetic", "genome")

corpus = tm\_map(corpus, removeWords, skipwords)

# term-document matrix

tdm = TermDocumentMatrix(corpus)

# convert tdm to matrix

m = as.matrix(tdm)

Matrixed word cloud seems legit. No small words unconnected to main topic

-http://wordnet.princeton.edu/



Bank List:

-See how they brand themselves (what do customers face)?

-Word list associated with each one

-Sometimes multiple Institutions (NAME) under one Bank Holding Company (NAMEHCR)

-Banks sometimes have a separate regional entity (ex. Bank of America California)

-Multiple entities based on type/class (Morgan Stanley Bank vs. Morgan Stanley Private Bank)

-Multiple entities as a result of a merger (JP Morgan Chase Bank vs. Chase Bank USA)

-Truly unique brands under a holding company (Whitney Bank under Hancock Holding Company)

-Common abbreviations and other name shortenings (BofA, Citi)

-Start with top 15, 10 more that are RBLC members that are different sizes

-Results: Big banks, esp. their Help/Ask twitter handles, people tweet at with both general and specific complaints & compliments

-With their main handles, or by searching by keywords, largely produces news results, business/economic predictions, giving out personal finance tips, and corporate sponsorships

-Sometimes people tweet complaints at a main handle, but generally these are redirected towards its customer service handle

-Smaller banks with only one Twitter handle has a mix of the two types, but more of the latter. Smaller banks have much fewer followers, which means a much lower volume as well

-Conclusion: Using Help/Ask handles might not be as proprietary as taking the entirety of what is said about financial institutions, but is the most focused on the results we are trying to get

-Using a keyword search is still possible, but would need an additional qualifier. We would need to use linguistical functions to determine what constitutes consumer experience vs. unrelated news and other comments about the institution (difficult)

**4. Look into R sentiment package**

Naïve Bayes

-Carlo Strapparava and Alessandro Valitutti’s emotions lexicon (emotions)

-http://www.cse.unt.edu/~rada/affectivetext/ (link not working)

-Janyce Wiebe’s subjectivity lexicon (polarity)

-<http://mpqa.cs.pitt.edu/lexicons/subj_lexicon/>

-http://people.cs.pitt.edu/~wiebe/pubs/papers/emnlp05polarity.pdf

-Starts with a lexicon of positive and negative words and phrases.

-A priori: out of context, does it evoke something positive or negative

-Negations (different types) reverse polarity

-Modality (proposition is asserted to be real or not real

-Word sense (word part of a phrase or by itself)

-Syntax

-Annotation scheme for marking contextual polarity of subjective expressions (positive, negative, both, neutral). Expressions tagged as emotions, evaluations, stances. Expressions then evaluated in context of the sentence

-Lexicon of 8,000 subjectivity clues from (Riloff and Wiebe, 2003)

-Marked strongly subjective or weakly subjective based on number of contexts in which they were subjective, and marked by polarity (a priori)

-Test: assign same clues polarity based on the polarity of expressions they appeared in (expressions were manually marked/annotated earlier). Clue marked as ‘Both’ if appeared in at least one positive and one negative expression

-Neutral Polar Classifier

-Word features: Word context (word itself, previous word, following word); prior polarity (positive, negative, both, neutral); reliability class (strongsubj or weaksubj)

-Modification features (binary relationships): Noun preceded by adjective; preceded by adverb other than not; preceded by intensifier; word itself is an intensifier; modifies/is modified by strongsubj/weaksubj (follows dependency tree)

-Structure features (binary features): determined by climbing up the dependency tree, looking for particular relationships words or patterns. In subject (subj relationship); in copular (true if in subject is false and if a node along the path is both a main verb and a copular verb); in passive (true if passive verb pattern is found)

-Sentence features: look for strongsubj and weaksubj clues in current, previous, and next sentences; binary features to indicate whether the sentence contains pronoun, cardinal number, and a modal other than *will*

-Document feature: topic of the document

-Voter procedure/algorithm (either)

*-Who has used it and what are they saying about it*

Notes on APIs:

[https://support.twitter.com/articles/160385-twitter-api-limits#](https://support.twitter.com/articles/160385-twitter-api-limits)

Next Steps (given body of 1,800 tweets):

-Perform sentiment analysis (either R sentiment package or Breen’s approach)

-Check to see if automatic ratings match up with manual judgment

-See if there is any bias, or if mistakes are random

-Find what issues are being talked about

-Categorize by issue or service (get clarification on this)

-Keyword searches?

Link up most frequent words to relevant banking words/issues