

NBTS SurveyMonkey Analysis

Amanda Bates

```
{r setup, include = FALSE, message = FALSE, warning = FALSE, cache = FALSE} chooseCRANmirror(graphics = FALSE, ind = 1) knitr::opts_chunk$set(echo = TRUE) Import data
```

```
# Import csv file
why <- read.csv('~/.MSDS692/Why_overall.csv', header = FALSE, stringsAsFactors = FALSE)
# Check the data
head(why)
```

```
##
## 1
## 2
## 3
## 4
## 5 I was not offered any experimental treatment, but instead was given the opportunity to be treated
## 6
```

Create corpus

```
# Load library
library(tm)
```

```
## Warning: package 'tm' was built under R version 3.4.3
```

```
## Loading required package: NLP
```

```
# Create corpus
why.corpus <- Corpus(VectorSource(why$V1))
```

Check the corpus

```
# Review corpus
inspect(why.corpus)
```

```
## <<SimpleCorpus>>
## Metadata: corpus specific: 1, document level (indexed): 0
## Content: documents: 21
##
## [1] Epidemiological studies only, not treatment
## [2] It was relayed to the bone paste in surgery. So not considered a treatment option.
## [3] Inoperable-Chemo & radiation would kill mutated gene causing cancer to become aggressive
## [4] my tumor was completely removed, no further treatment was needed. maybe I shouldn't be answering
## [5] I was not offered any experimental treatment, but instead was given the opportunity to be treated
## [6] It was through BrianScienceFoundation (Yale) study
## [7] I enrolled my wife in a clinical trial when standard of care treatment didn't work. She was rejected
## [8] enrolled but after dissection of tumor there was not enough left for the clinical trial
## [9] We realized the limitations of standard treatment nor did the first surgeon remove more the 50%
## [10] It was in the same class of drugs used to treat Jimmy Carter and at the same institution- sounds
## [11] We were desperate
## [12] He was terminal at diagnosis. Trials were the only hope.
## [13] She wanted to try something that might help her
## [14] It gave us hope that it would work
## [15] We wanted to maximize his odds of survival
## [16] Our cousin, a doctor at the hospital associated with the clinical trial, was instrumental in helping
```

```
## [17] My husband's decision, he wanted to try it.
## [18] She was being observed at NIH and they asked.
## [19] It was in the safety mode and he felt that by participating he could eventually help others
## [20] It was only trial he could get into at the time that had no restrictions as far as previous trea
## [21] He did hyperbaric radiation which was new. Don't know if it counts
```

Data cleaning Transform corpus

```
# Remove whitespace
why.corpus.cln <- tm_map(why.corpus, stripWhitespace)
# Make lowercase
why.corpus.cln <- tm_map(why.corpus.cln, content_transformer(tolower))
# Remove stopwords
why.corpus.cln <- tm_map(why.corpus.cln, removeWords, stopwords("english"))
# Remove punctuation
why.corpus.cln <- tm_map(why.corpus.cln, content_transformer(removePunctuation))

# Check corpus
inspect(why.corpus.cln)
```

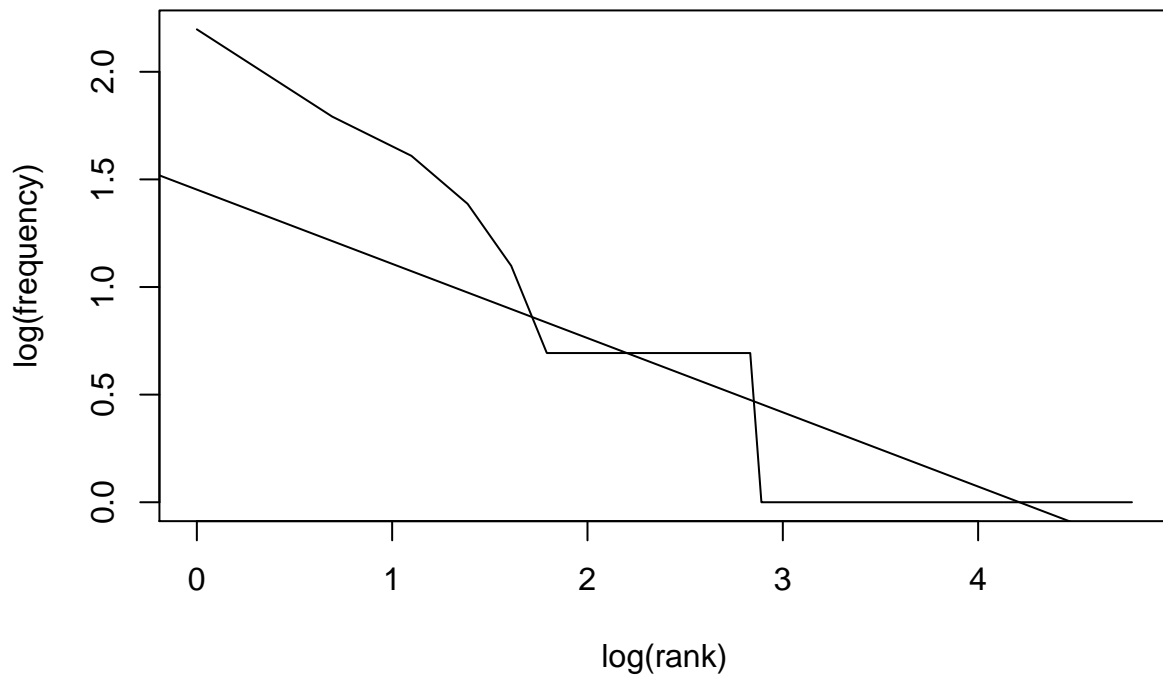
```
## <<SimpleCorpus>>
## Metadata: corpus specific: 1, document level (indexed): 0
## Content: documents: 21
##
## [1] epidemiological studies treatment
## [2] relayed bone paste surgery considered treatment option
## [3] inoperablechemo radiation kill mutated gene causing cancer become aggressive
## [4] tumor completely removed treatment needed maybe answering survey
## [5] offered experimental treatment instead given opportunity treated nih continue tra
## [6] briansciencefoundation yale study
## [7] enrolled wife clinical trial standard care treatment work rejected clinical trial
## [8] enrolled dissection tumor enough left clinical trial
## [9] realized limitations standard treatment first surgeon remove 50 tumor
## [10] class drugs used treat jimmy carter institution sounded promising
## [11] desperate
## [12] terminal diagnosis trials hope
## [13] wanted try something might help
## [14] gave us hope work
## [15] wanted maximize odds survival
## [16] cousin doctor hospital associated clinical trial instrumental helping choose treatment
## [17] husbands decision wanted try
## [18] observed nih asked
## [19] safety mode felt participating eventually help others
## [20] trial get time restrictions far previous treatments avastin
## [21] hyperbaric radiation new know counts
```

Data Exploration Create a DTM and TDM

```
# Create document term matrix
why.dtm <- DocumentTermMatrix(why.corpus.cln)
# Create term document matrix
why.tdm <- TermDocumentMatrix(why.corpus.cln)
```

Create Zipf's plot

```
# Create Zipf's plot
Zipf_plot(why.tdm)
```



```
## (Intercept)          x
##  1.4526350 -0.3449341
```

Find terms that appear at least 5 times

```
# Find terms that occur at least 5 times
findFreqTerms(why.dtm, 5)
```

```
## [1] "treatment" "tumor"      "trial"
```

Data Visualization Create a wordcloud

```
# Load required library
library(wordcloud)
```

```
## Warning: package 'wordcloud' was built under R version 3.4.2
```

```
## Loading required package: RColorBrewer
```

```
# Create wordcloud with the most commonly used words larger and in the center
wordcloud(why.corpus.cln, random.order = FALSE)
```

clinical tumor treatment trial wanted

Text Clustering # Create lda model with 4 topics

```
# Load required library
library(topicmodels)
```

```
## Warning: package 'topicmodels' was built under R version 3.4.2
```

```
# Create LDA model with 4 topics
why.lda <- LDA(why.dtm, k = 4)
# Review 7 terms associated with each topic
term <- terms(why.lda, 7)
(term <- apply(term, MARGIN = 2, paste, collapse = ", "))
```

```
##                                     Topic 1
## "treatment, help, try, wanted, epidemiological, studies, radiation"
##                                     Topic 2
##      "treatment, trial, bone, considered, option, paste, relayed"
##                                     Topic 3
##      "treatment, tumor, addition, bank, best, brain, care"
##                                     Topic 4
##      "tumor, clinical, trial, treatment, enrolled, work, hope"
```

Sentiment analysis

```
# Load library
library(SentimentAnalysis)
```

```
## Warning: package 'SentimentAnalysis' was built under R version 3.4.3
```

```
##
```

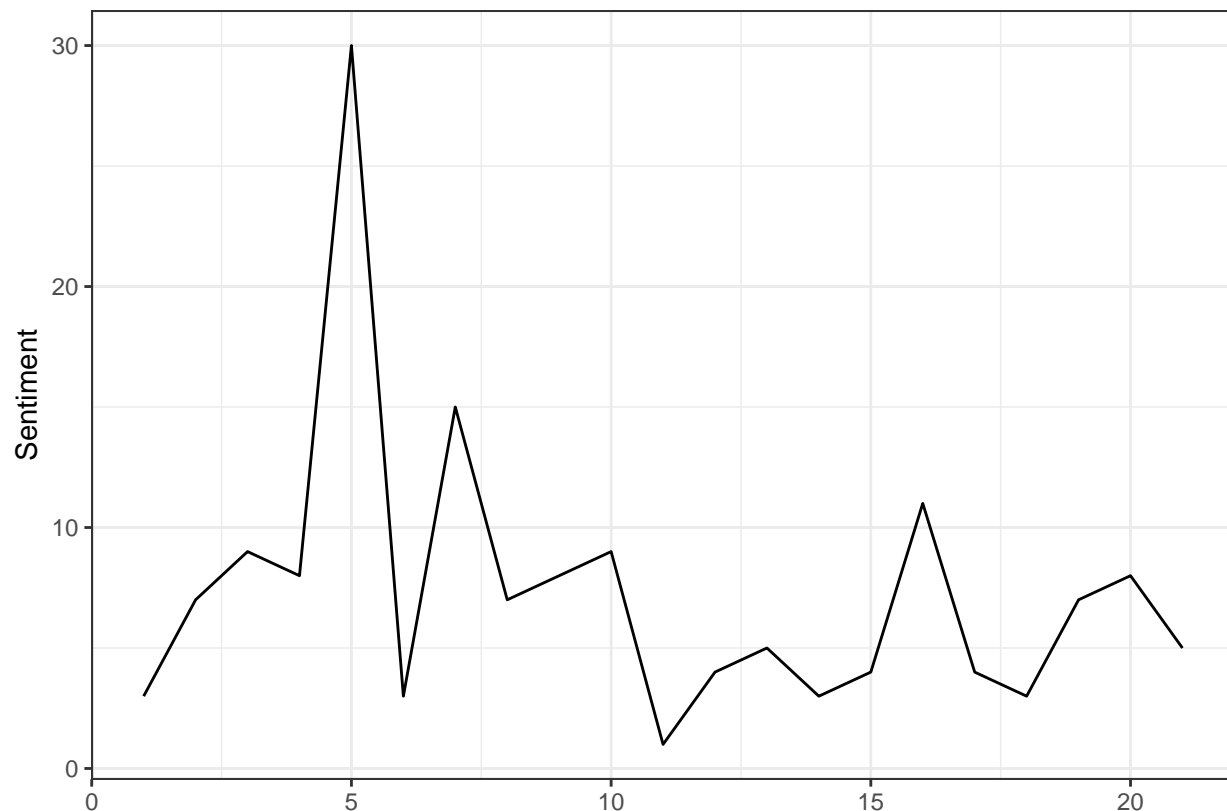
```
## Attaching package: 'SentimentAnalysis'

## The following object is masked from 'package:base':
##
##      write

# Perform sentiment analysis
why.sentiment <- analyzeSentiment(why.corpus.cln)
# Check the output of the sentiment analysis
head(why.sentiment)

##      WordCount SentimentGI NegativityGI PositivityGI SentimentHE NegativityHE
## 1           3  -0.3333333   0.3333333   0.0000000         0.0         0
## 2           7   0.2857143   0.0000000   0.2857143         0.0         0
## 3           9  -0.2222222   0.3333333   0.1111111         0.0         0
## 4           8   0.0000000   0.1250000   0.1250000         0.0         0
## 5          30   0.3000000   0.1000000   0.4000000         0.1         0
## 6           3  -0.3333333   0.3333333   0.0000000         0.0         0
##      PositivityHE SentimentLM NegativityLM PositivityLM RatioUncertaintyLM
## 1           0.0         0.0         0         0.0         0.000
## 2           0.0         0.0         0         0.0         0.000
## 3           0.0         0.0         0         0.0         0.000
## 4           0.0         0.0         0         0.0         0.125
## 5           0.1         0.1         0         0.1         0.000
## 6           0.0         0.0         0         0.0         0.000
##      SentimentQDAP NegativityQDAP PositivityQDAP
## 1      0.0000000      0.0000000      0.0000000
## 2      0.0000000      0.0000000      0.0000000
## 3     -0.3333333      0.3333333      0.0000000
## 4      0.0000000      0.0000000      0.0000000
## 5      0.2666667      0.0000000      0.2666667
## 6      0.0000000      0.0000000      0.0000000

# Visualize the results of the sentiment analysis
plotSentiment(why.sentiment)
```



```
# Get summary results from sentiment analysis
summary(why.sentiment)
```

```
##      WordCount      SentimentGI      NegativityGI      PositivityGI
## Min.   : 1.000      Min.   : -1.000000      Min.   : 0.000      Min.   : 0.0000
## 1st Qu.: 4.000      1st Qu.: -0.222222      1st Qu.: 0.000      1st Qu.: 0.0000
## Median : 7.000      Median : 0.000000      Median : 0.125      Median : 0.2000
## Mean   : 7.333      Mean   : 0.006315      Mean   : 0.181      Mean   : 0.1873
## 3rd Qu.: 8.000      3rd Qu.: 0.250000      3rd Qu.: 0.250      3rd Qu.: 0.2857
## Max.   :30.000      Max.   : 0.444444      Max.   : 1.000      Max.   : 0.5000
##      SentimentHE      NegativityHE      PositivityHE      SentimentLM
## Min.   :0.000000      Min.   : 0      Min.   : 0.000000      Min.   : -0.25000
## 1st Qu.:0.000000      1st Qu.: 0      1st Qu.: 0.000000      1st Qu.: 0.00000
## Median :0.000000      Median : 0      Median : 0.000000      Median : 0.00000
## Mean   :0.004762      Mean   : 0      Mean   : 0.004762      Mean   : -0.02262
## 3rd Qu.:0.000000      3rd Qu.: 0      3rd Qu.: 0.000000      3rd Qu.: 0.00000
## Max.   :0.100000      Max.   : 0      Max.   : 0.100000      Max.   : 0.10000
##      NegativityLM      PositivityLM      RatioUncertaintyLM
## Min.   :0.000000      Min.   : 0.000000      Min.   : 0.00000
## 1st Qu.:0.000000      1st Qu.: 0.000000      1st Qu.: 0.00000
## Median :0.000000      Median : 0.000000      Median : 0.00000
## Mean   :0.03056      Mean   : 0.007937      Mean   : 0.01548
## 3rd Qu.:0.000000      3rd Qu.: 0.000000      3rd Qu.: 0.00000
## Max.   :0.25000      Max.   : 0.100000      Max.   : 0.20000
##      SentimentQDAP      NegativityQDAP      PositivityQDAP
## Min.   : -1.00000      Min.   : 0.00000      Min.   : 0.00000
```

```
## 1st Qu.: 0.00000 1st Qu.:0.00000 1st Qu.:0.0000
## Median : 0.00000 Median :0.00000 Median :0.1333
## Mean : 0.05993 Mean :0.09048 Mean :0.1504
## 3rd Qu.: 0.25000 3rd Qu.:0.06667 3rd Qu.:0.2500
## Max. : 0.66667 Max. :1.00000 Max. :0.6667
```

Create n-grams

```
# Convert corpus to strings
why.string <- as.character(why.corpus.cln)[1]
# Separate words
why.words <- strsplit(why.string, " ", fixed = TRUE) [[1]]
library(NLP)
# create bigrams
why.bi <- vapply(ngrams(why.words, 2), paste, "", collapse = " ")
# Check bigrams
why.bi[1:3]
```

```
## [1] "c(\"epidemiological studies\" \"studies \"
## [3] " "
```

Text Clustering with n-grams

```
library(RWeka)
```

```
## Warning: package 'RWeka' was built under R version 3.4.2
```

```
# Create bigramTokenizer
BigramTokenizer <- function(x) NGramTokenizer(x, Weka_control(min = 2, max = 2))
# Create bigram DTM
why.bi.dtm <- DocumentTermMatrix(why.corpus.cln,
                                control = list(tokenize = BigramTokenizer))
# Create bigram LDA model
why.bigram.lda <- LDA(why.bi.dtm, k = 3)
# Review 7 terms associated with each topic
term.bi <- terms(why.bigram.lda, 7)
(term.bi <- apply(term.bi, MARGIN = 2, paste, collapse = ", "))
```

```
##                                     Topic 1
##               "trial, tumor, clinical, work, hope, try, wanted"
##                                     Topic 2
## "treatment, nih, epidemiological, studies, tumor, addition, bank"
##                                     Topic 3
## "treatment, tumor, clinical, trial, bone, considered, option"
```

K-means clustering

```
# Create new DTM with TF-IDF
why.tfidf <- DocumentTermMatrix(why.corpus.cln, control = list(weighting = weightTfIdf))
# Check dimensions of DTM
dim(why.tfidf)
```

```
## [1] 21 120
```

```
# Run k-means with 3 clusters
why.kmeans <- kmeans(why.tfidf, 3, iter.max = 100)
table(why.kmeans$cluster)
```

```
##
```

```
## 1 2 3
## 1 1 19
# Identify ten words associated with each cluster
for (i in 1:3) {
  cat(paste("Cluster ", i, ": ", sep = ""))
  s <- sort(why.kmeans$centers[i, ], decreasing = TRUE)
  cat(names(s)[1:10], "\n")
}
```

```
## Cluster 1: desperate epidemiological studies treatment bone considered option paste relayed surgery
## Cluster 2: briansciencefoundation study yale epidemiological studies treatment bone considered option
## Cluster 3: hope wanted treatment try epidemiological studies gave asked observed work
```

Additional question analysis Import data

```
# Import csv file
whynot <- read.csv('~/.MSDS692/not_overall.csv', header = FALSE, stringsAsFactors = FALSE)
# Check the data
head(whynot)
```

```
##
## 1 my treatment plan appeared sufficient V1
## 2 My tumor was 100 % gross resection as was my recurrence so I haven't needed a clinical trial.
## 3 we did not know of any
## 4 Prognosis seemed good without participating
## 5 Do not need trial
## 6 There were no ongoing trials in the U.S. regarding my diagnosis.
```

Create corpus

```
# Create corpus
whynot.corpus <- Corpus(VectorSource(whynot$V1))
# Review corpus
inspect(whynot.corpus)
```

```
## <<SimpleCorpus>>
## Metadata: corpus specific: 1, document level (indexed): 0
## Content: documents: 206
##
## [1] my treatment plan appeared sufficient
## [2] My tumor was 100 % gross resection as was my recurrence so I haven't needed a clinical trial.
## [3] we did not know of any
## [4] Prognosis seemed good without participating
## [5] Do not need trial
## [6] There were no ongoing trials in the U.S. regarding my diagnosis.
## [7] After surgical resection and proton beam radiation my prognosis was very positive
## [8] I wanted to give the current standard treatments a chance first before relying on clinical trial
## [9] I don't think my condition (nonmalignant tumor) has any clinical trials
## [10] My tumor was straightforward
## [11] appropriate trial not available
## [12] My current treatment is working
## [13] I was not eligible as I did not fit the trial criteria
## [14] Would have to board one dog and fight to have my service dog accompany me.
## [15] Not apply
## [16] There wasn't any time nor were there any clinical trials available for my type of cancer
## [17] Had TUMOUR removed
```


[18] At this time no treatment is recommended as long as tumor is stable
 ## [19] My tumor isn't life-threatening and trusted my doctor course of action
 ## [20] I chose standard treatment & care.
 ## [21] I could not find a clinical trial that matched the type of brain tumor I have
 ## [22] it is not a scare yet
 ## [23] Not sure one exists and probably wouldn't take it
 ## [24] The biopsy did not capture enough tumor tissue to determine the type and origin. When the surg
 ## [25] I'm military....So, I guess it IS trial ...of sorts. lol
 ## [26] I'm military....So, I guess it IS trial ...of sorts. lol
 ## [27] Most trials take too long to benefit someone with relatively short life span
 ## [28] I was at UCLA and saw the head of neurosurgery and never recommended
 ## [29] My brain tumor surgery was in 1998
 ## [30] I had successful recovery
 ## [31] My tumor was removed & left no deficits. After 18 years of follow up, it has not returned. I ha
 ## [32] Didn't feel like I needed a clinical trial; other options to consider first
 ## [33] Surgery was highly recommended by my doctors and it was beneficial in removing the tumor with
 ## [34] I was fully unconscious due to 'giant', 'ancient' tumor
 ## [35] I didn't want to
 ## [36] There was no trial offered... a treatment plan had been established for my type of tumor.
 ## [37] Additional treatment not needed at this time
 ## [38] I did participate in a clinical procedure during surgery involving pathological exam to make su
 ## [39] I feel stable still, so no need.
 ## [40] I was successfully treated with 2 surgeries and radiation
 ## [41] My team felt confident that a successful surgery was the best step. I have MRIs every 6 months
 ## [42] I was told my meningioma was not malignant, didn't need anything other than surgery
 ## [43] My parents were in failing health (my mother was dying and my father's health was spiraling do
 ## [44] The results of the clinical trial for the vaccine I was scheduled to receive showed that it c
 ## [45] Wasn't offered one, and the treatment I was given seems to have worked
 ## [46] There are no clinical trials available based on my current condition. Trials seem to be geared
 ## [47] Education ME, willing to help the future patients with GBM
 ## [48] I am still completing standard treatment
 ## [49] I am doing fine after surgery and radiation
 ## [50] Wasn't informed
 ## [51] My tumor was so big I would have died if I waited
 ## [52] My doctor did not want me to wait till trial opened up
 ## [53] X
 ## [54] No time- surgery to remove was shortly after diagnosis.
 ## [55] no need for a clinical trial
 ## [56] At the time of diagnosis he was 95% sure it was a benign meningioma. Surgery diagnosed atypical
 ## [57] No time. I had my surgery 4 days after being diagnosed.
 ## [58] I don't think I would have qualified as I had a meningioma.
 ## [59] Not applicable to my situation
 ## [60] I have not yet reached a point since surgery that my doctor feels we need to review further tr
 ## [61] Current treatment plan is working
 ## [62] The standard protocol for my tumor has been working. We've discussed a clinical trial should w
 ## [63] Not necessary - Benign tumor was surgically removed
 ## [64] Treatment began so quickly after I was diagnosed and clinical trials weren't brought up by my c
 ## [65] I have a low grade glioma and do not qualify; trials found are for GBMs
 ## [66] Surgery was recommended immediately and tumor was removed.
 ## [67] At this point, my treatment was successful
 ## [68] I had 100% resection of the tumor
 ## [69] I didn't want to. Participate n a trial at that time
 ## [70] There isn't one for what I have
 ## [71] Doctor said none apply to my type of tumor currently.

[72] Not aware of all clinical trials that may be best options
 ## [73] Tumor Review Board recommended standard treatment
 ## [74] Trial was stopped before I could decidedly about joining
 ## [75] Would not consider a clinical trial
 ## [76] surgery sufficient for tumor type at time
 ## [77] My tumor was large and aggressive and had to have severely as quickly as possible.
 ## [78] My tumor has not recurred. I am a patient at Duke.
 ## [79] Traditional surgery was expected to be successful and was
 ## [80] I have been remission for 10 years
 ## [81] The type of tumor I have does not qualify for clinical trials; surgical intervention is the on
 ## [82] Glioblastoma is stable with Avastin now -- did not qualify
 ## [83] I have been told I am stable and likely benign. Even though I have symptoms, the risks likely c
 ## [84] I researched trials when first diagnosed and my oncologist knew nothing about the. My current m
 ## [85] I did ask the doctor, but he said there weren't any for my atypical meningioma.
 ## [86] My doctor said my tumor is too large (9cm across both frontal lobes).
 ## [87] I don't feel it necessary at this point. SOC and Optune device has kept my tumor stable.
 ## [88] I felt the options provided by my doctor were the right decisions
 ## [89] My doctor was able to 100% remove my tumor surgically
 ## [90] I think I was treatable and offered my results and blood if they needed it.
 ## [91] I had a very straight forward case that had well documented successful treatment options
 ## [92] there are few clinical trials available for meningioma patients, probably zero for grade 1 men
 ## [93] My treatment plan was always straightforward and clear.. and was successful overall.
 ## [94] Not needed at this time
 ## [95] My last surgery was in 09. I don't need a clinical trial now
 ## [96] There were no clinical studies that were relevant at diagnosis or recurrence.
 ## [97] My gene codes were sent to a company that does clinical trials and I was told I would be conta
 ## [98] Clinical trials for my recurrence are TBD pending biopsy results and review/application of tri
 ## [99] wasn't time, had to have surgery next day
 ## [100] in both instances (first and recurrent), my doctor advised that first choice therapy (surgery :
 ## [101] Had huge faith in my doctor.
 ## [102] Needed treatment immediately
 ## [103] My treatment plan was always straightforward and clear.. and was successful overall.
 ## [104] I do not qualify for clinical trials at this moment in time.
 ## [105] Thus far, radiation and chemotherapy have worked.
 ## [106] I am personally looking into them myself. This survey should go to all oncologists AND patien
 ## [107] Surgical removal of the tumor was deemed the most effective option in my case
 ## [108] Oncologist said there were no clinical trials.
 ## [109] There were no Clinical trials
 ## [110] my husband only lived 9 months from day we found out he had 5 brain tumors, we did radiation,
 ## [111] No appropriate trials for him. Died in 2012
 ## [112] The clinical trial available, Temodar for Glioblastoma was already being used for Astrocytomas
 ## [113] At the time it was presented, it was not made clear that the clinical trial was in ADDITION to
 ## [114] Patient's condition deteriorated rapidly under standard of care and his condition was too poor
 ## [115] My father passed away less that 2 years after diagnosis and after much unnecessary sufferings.
 ## [116] The time from diagnosis till death was only seventeen days.
 ## [117] Doctor says tumor recurrence has to occur before eligible for trial.
 ## [118] He had resection of R Temporal Lobe which led to diagnosis of GBM. He then went to MD Anderson
 ## [119] Her platelets were too low to ever qualify for a trial
 ## [120] He did not need a clinical trial as he responded to normal treatment
 ## [121] Informed that standard Protocol treatment had to be completed before any trials were an option
 ## [122] Prime objective was to cause dormancy of a brain stem tumor...achieved with chemo drugs.
 ## [123] Was told by the doctor she was receiving the best treatment for her condition and clinical tri
 ## [124] Doctors told us we had to research them ourselves nearly a year into diagnosis - they should've
 ## [125] Stroke was too devastating and second tumor not accessible

[126] The clinical trial we were informed about is 2 hours from our home, during winter with travel
 ## [127] Surgery almost 30 years ago and tumor totally removed followed by radiation therapy. No trial
 ## [128] selected the standard treatment (radiation and chemo) for his tumor type which was the recommen
 ## [129] Haven't gotten there yet
 ## [130] We were told not at that point yet, then it was a last option, his scores were not high enough
 ## [131] My husband did not have the right markers for 2 trials that we thought he would be enrolled.
 ## [132] I don't believe that it was ever suggested.
 ## [133] Tumor was too aggressive. Time from diagnosis to death was 20 days
 ## [134] The FDA pulled a trial my son was going to do just before he was supposed to stop. We decided
 ## [135] Too far along when the tumor was discovered.
 ## [136] mostly did not qualify for any because he had been treated with Avastin, early on.
 ## [137] Legal forms presented were very difficult to understand and frightening
 ## [138] he passed away the day of diagnosis
 ## [139] Deteriorated whilst waiting for 2nd opinion after recurrence. Never got the chance to find out
 ## [140] Due to the aggressiveness of the tumor a beneficial clinical trial for her was not available.
 ## [141] We wanted to move quickly to treat the tumor with the treatment plan recommended by his doctor
 ## [142] doctor said no trials till has reoccurrence
 ## [143] Access to clinical trials was almost non existent in Australia when my husband was diagnosed in
 ## [144] she felt she gave it her best shot and at that time a clinical wasn't going to save her life
 ## [145] Had been previously misdiagnosed with MS and doctors are concerned it may have not been a comp
 ## [146] Clinical trials were not mentioned to us by any Doctor.
 ## [147] CHOP said the Thalamic tumor was inoperable. We explored 7 specialized facilities including M
 ## [148] We wanted my son to have the surgery
 ## [149] Didn't have enough time
 ## [150] Simply too hard to find information and sort it out.
 ## [151] Could not participate because had already done the gamma knife
 ## [152] Did not need them
 ## [153] Tumor was too advanced - no time to enroll in a trial
 ## [154] three standard of care options given first, then clinical trials discussed
 ## [155] Clinical trial was suggested after standard treatment was no longer effective. After 2.5 years
 ## [156] New procedure recommended involving laser to kill the tumor
 ## [157] We did not ask, nor were the options offered. His diagnosis was terminal and palliative care
 ## [158] He was told that there could be clinical trials if there was a recurrence. No details on how
 ## [159] Suffered seizures and was unable to participate
 ## [160] Was evaluated at NIH and was about to enter a train when recurrence was found
 ## [161] Standard Care treatment not completed.
 ## [162] Tumor caused hemorrhagic stroke and significant detriment
 ## [163] Given that there was no biopsy due to the urgency to begin treatment, clinical trials were not
 ## [164] Didn't have any clinical trials for him at the time.
 ## [165] I don't know if it was discussed; I did not hear about any such offer.
 ## [166] Participated in a natural history study at NIH but not a clinical trial of medication or treat
 ## [167] we are still waiting to find a good one
 ## [168] He had been treated for prostate cancer (not related) and did not qualify
 ## [169] Given my mother's age and the location and type of tumor, it was expressed that it would not p
 ## [170] He passed away before we could get a second opinion and find out about clinical trial
 ## [171] QUICK SURGERY WAS THE ONLY RECOMMENDED TREATMENT
 ## [172] it was never brought up
 ## [173] Not available
 ## [174] By the time we were informed about them, he was too sick to qualify for anything
 ## [175] Also did not have genetic mutation
 ## [176] Afraid of not receiving the gold standard of treatment and receiving only the new treatment tha
 ## [177] Still waiting for final treatment plan from oncologist
 ## [178] None that were providing better treatment than standard.
 ## [179] Not an option at time of initial treatment and recurrence.

```

## [180] We are considering a clinical trial, the uw has not offered any
## [181] At no time were any clinical trials discussed between doctor and patient
## [182] Standard treatment at the time offered generally good results
## [183] Standard treatment was successful
## [184] MUM NEVER KNEW SHE HAD A BRAIN TUMOUR
## [185] We chose a non-trial treatment that was more aggressive than that offered in a clinical trial
## [186] Tumor was so rare, a Clinical trial couldn't be specific enough
## [187] He died in 1971, so there were no clinical trials then
## [188] We were just starting to consider clinical trials as standard treatment had worked for 12 years
## [189] The clinical trial was delayed due to not enough staff and my brother in law then had another
## [190] My wife responded to standard treatment. She just had a second recurrence in august 2016 and w
## [191] Quality of life more important than quantity to son-concerns that clinical trial may potential
## [192] we were offered two trials but both were very involved and would have to go to the city where
## [193] My brother thought he would survive, he had complete faith in his doctors which having met the
## [194] She was not eligible for clinical trials. I work in cancer research at one of the top cancer c
## [195] The doctors did not suggest any
## [196] Don't know
## [197] he thought it was too aggressive at the time
## [198] Patient is already undergoing standard treatment that is working.
## [199] Decided to finish college before starting treatment
## [200] Most of the promising clinical trials would not accept a patient who had started radiation the
## [201] It all happened in such a short period of time.
## [202] Kaiser did not bother to give any options. When asked about options Kaiser said they did offer
## [203] The total rudeness of doctor at Duke
## [204] Wanted to enjoy quality of life for the remaining 6 months expected
## [205] Was not presented as an option.
## [206] They said he was too old

```

Data cleaning Transform corpus

```

# Remove whitespace
whynot.corpus.cln <- tm_map(whynot.corpus, stripWhitespace)
# Make lowercase
whynot.corpus.cln <- tm_map(whynot.corpus.cln, content_transformer(tolower))
# Remove stopwords
whynot.corpus.cln <- tm_map(whynot.corpus.cln, removeWords, stopwords("english"))
# Remove punctuation
whynot.corpus.cln <- tm_map(whynot.corpus.cln, content_transformer(removePunctuation))
# Check corpus
inspect(whynot.corpus.cln)

```

```

## <<SimpleCorpus>>
## Metadata: corpus specific: 1, document level (indexed): 0
## Content: documents: 206
##
## [1] treatment plan appeared sufficient
## [2] tumor 100 gross resection recurrence needed clinical trial
## [3] know
## [4] prognosis seemed good without participating
## [5] need trial
## [6] ongoing trials us regarding diagnosis
## [7] surgical resection proton beam radiation prognosis positive
## [8] wanted give current standard treatments chance first relying clinical trials
## [9] think condition nonmalignant tumor clinical trials
## [10] tumor straightforward

```

[11] appropriate trial available
 ## [12] current treatment working
 ## [13] eligible fit trial criteria
 ## [14] board one dog fight service dog accompany
 ## [15] apply
 ## [16] time clinical trials available type cancer
 ## [17] tumour removed
 ## [18] time treatment recommended long tumor stable
 ## [19] tumor lifethreatening trusted doctor course action
 ## [20] chose standard treatment care
 ## [21] find clinical trial matched type brain tumor
 ## [22] scare yet
 ## [23] sure one exists probably take
 ## [24] biopsy capture enough tumor tissue determine type origin surgery finished tumor r
 ## [25] military guess trial sorts lol
 ## [26] military guess trial sorts lol
 ## [27] trials take long benefit someone relatively short life span
 ## [28] ucla saw head neurosurgery never recommended
 ## [29] brain tumor surgery 1998
 ## [30] successful recovery
 ## [31] tumor removed left deficits 18 years follow returned needed treatment actively
 ## [32] feel like needed clinical trial options consider first
 ## [33] surgery highly recommended doctors beneficial removing tumor regrowth point treat
 ## [34] fully unconscious due giant ancient tumor
 ## [35] want
 ## [36] trial offered treatment plan established type tumor
 ## [37] additional treatment needed time
 ## [38] participate clinical procedure surgery involving pathological exam make sure got clean r
 ## [39] feel stable still need
 ## [40] successfully treated 2 surgeries radiation
 ## [41] team felt confident successful surgery best step mris every 6 months
 ## [42] told meningioma malignant need anything surgery
 ## [43] parents failing health mother dying fathers health spiraling downward find housing v
 ## [44] results clinical trial vaccine scheduled receive showed provide benefit
 ## [45] offered one treatment given seems worked
 ## [46] clinical trials available based current condition trials seem geared towards respondi
 ## [47] education willing help future patients gbm
 ## [48] still completing standard treatment
 ## [49] fine surgery radiation
 ## [50] informed
 ## [51] tumor big died waited
 ## [52] doctor want wait till trial opened
 ## [53] x
 ## [54] time surgery remove shortly diagnosis
 ## [55] need clinical trial
 ## [56] time diagnosis 95 sure benign meningioma surgery diagnosed atypical grade ii clinical
 ## [57] time surgery 4 days diagnosed
 ## [58] think qualified meningioma
 ## [59] applicable situation
 ## [60] yet reached point since surgery doctor feels need review treatment options yet
 ## [61] current treatment plan working
 ## [62] standard protocol tumor working discussed clinical trial run options yet
 ## [63] necessary benign tumor surgically removed
 ## [64] treatment began quickly diagnosed clinical trials brought doctors sure deemed c

[65] low grade glioma qualify trials found gbms
 ## [66] surgery recommended immediately tumor removed
 ## [67] point treatment successful
 ## [68] 100 resection tumor
 ## [69] want participate n trial time
 ## [70] one
 ## [71] doctor said none apply type tumor currently
 ## [72] aware clinical trials may best options
 ## [73] tumor review board recommended standard treatment
 ## [74] trial stopped decidedly joining
 ## [75] consider clinical trial
 ## [76] surgery sufficient tumor type time
 ## [77] tumor large aggressive severely quickly possible
 ## [78] tumor recurred patient duke
 ## [79] traditional surgery expected successful
 ## [80] remission 10 years
 ## [81] type tumor qualify clinical trials surgical intervention treatment
 ## [82] glioblastoma stable avastin now qualify
 ## [83] told stable likely benign even though symptoms risks likely outweigh benefits ti
 ## [84] researched trials first diagnosed oncologist knew nothing current neuro oncologist seen
 ## [85] ask doctor said atypical meningioma
 ## [86] doctor said tumor large 9cm across frontal lobes
 ## [87] feel necessary point soc optune device kept tumor stable
 ## [88] felt options provided doctor right decisions
 ## [89] doctor able 100 remove tumor surgically
 ## [90] think treatable offered results blood needed
 ## [91] straight forward case well documented successful treatment options
 ## [92] clinical trials available meningioma patients probably zero grade 1 meningioma patients b
 ## [93] treatment plan always straightforward clear successful overall
 ## [94] needed time
 ## [95] last surgery 09 need clinical trial now
 ## [96] clinical studies relevant diagnosis recurrence
 ## [97] gene codes sent company clinical trials told contacted type codes came clinic
 ## [98] clinical trials recurrence tbd pending biopsy results reviewapplication trials diagnosis
 ## [99] time surgery next day
 ## [100] instances first recurrent doctor advised first choice therapy surgery first time radiati
 ## [101] huge faith doctor
 ## [102] needed treatment immediately
 ## [103] treatment plan always straightforward clear successful overall
 ## [104] qualify clinical trials moment time
 ## [105] thus far radiation chemotherapy worked
 ## [106] personally looking survey go oncologists patients well second opinionsthanks a
 ## [107] surgical removal tumor deemed effective option case
 ## [108] oncologist said clinical trials
 ## [109] clinical trials
 ## [110] husband lived 9 months day found 5 brain tumors radiation chemo avastin nova care
 ## [111] appropriate trials died 2012
 ## [112] clinical trial available temodar glioblastoma already used astrocytomas unnecessary j
 ## [113] time presented made clear clinical trial addition standard care temodarradiation
 ## [114] patients condition deteriorated rapidly standard care condition poor qualify trial s
 ## [115] father passed away less 2 years diagnosis much unnecessary sufferings inadequate medical
 ## [116] time diagnosis till death seventeen days
 ## [117] doctor says tumor recurrence occur eligible trial
 ## [118] resection r temporal lobe led diagnosis gbm went md anderson radiation chemo neve

[119] platelets low ever qualify trial
 ## [120] need clinical trial responded normal treatment
 ## [121] informed standard protocol treatment completed trials option
 ## [122] prime objective cause dormancy brain stem tumor achieved chemo drugs
 ## [123] told doctor receiving best treatment condition clinical trial apply
 ## [124] doctors told us research nearly year diagnosis ve told us look start found one 1
 ## [125] stroke devastating second tumor accessible
 ## [126] clinical trial informed 2 hours home winter travel snow belt area new york state
 ## [127] surgery almost 30 years ago tumor totally removed followed radiation therapy trial mentioned
 ## [128] selected standard treatment radiation chemo tumor type recommendation primary physician
 ## [129] gotten yet
 ## [130] told point yet last option scores high enough since put avastin start qualify
 ## [131] husband right markers 2 trials thought enrolled platelet count low certain drugs
 ## [132] believe ever suggested
 ## [133] tumor aggressive time diagnosis death 20 days
 ## [134] fda pulled trial son going just supposed stop decided final clinical trial sw
 ## [135] far along tumor discovered
 ## [136] mostly qualify treated avastin early
 ## [137] legal forms presented difficult understand frightening
 ## [138] passed away day diagnosis
 ## [139] deteriorated whilst waiting 2nd opinion recurrence never got chance find trials might
 ## [140] due aggressiveness tumor beneficial clinical trial available death trial ava
 ## [141] wanted move quickly treat tumor treatment plan recommended doctor brain tumor board
 ## [142] doctor said trials till reoccurrence
 ## [143] access clinical trials almost non existent australia husband diagnosed 2000
 ## [144] felt gave best shot time clinical going save life
 ## [145] previously misdiagnosed ms doctors concerned may complete misdiagnosis afraid give
 ## [146] clinical trials mentioned us doctor
 ## [147] chop said thalamic tumor inoperable explored 7 specialized facilities including mayo john h
 ## [148] wanted son surgery
 ## [149] enough time
 ## [150] simply hard find information sort
 ## [151] participate already done gamma knife
 ## [152] need
 ## [153] tumor advanced time enroll trial
 ## [154] three standard care options given first clinical trials discussed
 ## [155] clinical trial suggested standard treatment longer effective 25 years conditions worsened
 ## [156] new procedure recommended involving laser kill tumor
 ## [157] ask options offered diagnosis terminal palliative care
 ## [158] told clinical trials recurrence details find one help
 ## [159] suffered seizures unable participate
 ## [160] evaluated nih enter train recurrence found
 ## [161] standard care treatment completed
 ## [162] tumor caused hemorrhagic stroke significant detriment
 ## [163] given biopsy due urgency begin treatment clinical trials option
 ## [164] clinical trials time
 ## [165] know discussed hear offer
 ## [166] participated natural history study nih clinical trial medication treatment
 ## [167] still waiting find good one
 ## [168] treated prostate cancer related qualify
 ## [169] given mothers age location type tumor expressed provide extended life good quality
 ## [170] passed away get second opinion find clinical trial
 ## [171] quick surgery recommended treatment
 ## [172] never brought

```

## [173] available
## [174] time informed sick qualify anything
## [175] also genetic mutation
## [176] afraid receiving gold standard treatment receiving new treatment evidence effective
## [177] still waiting final treatment plan oncologist
## [178] none providing better treatment standard
## [179] option time initial treatment recurrence
## [180] considering clinical trial uw offered
## [181] time clinical trials discussed doctor patient
## [182] standard treatment time offered generally good results
## [183] standard treatment successful
## [184] mum never knew brain tumour
## [185] chose nontrial treatment aggressive offered clinical trial treatment center
## [186] tumor rare clinical trial specific enough
## [187] died 1971 clinical trials
## [188] just starting consider clinical trials standard treatment worked 12 years
## [189] clinical trial delayed due enough staff brother law another tumor recurrence delay
## [190] wife responded standard treatment just second recurrence august 2016 eligible clinical
## [191] quality life important quantity sonconcerns clinical trial may potentially diminish quality
## [192] offered two trials involved go city treated couple times week 150 miles round
## [193] brother thought survive complete faith doctors met can understand
## [194] eligible clinical trials work cancer research one top cancer centers country enrollment
## [195] doctors suggest
## [196] know
## [197] thought aggressive time
## [198] patient already undergoing standard treatment working
## [199] decided finish college starting treatment
## [200] promising clinical trials accept patient started radiation therapy
## [201] happened short period time
## [202] kaiser bother give options asked options kaiser said offer clinical trials
## [203] total rudeness doctor duke
## [204] wanted enjoy quality life remaining 6 months expected
## [205] presented option
## [206] said old

```

Data Exploration Create a DTM and TDM

```

# Create document term matrix
whynot.dtm <- DocumentTermMatrix(whynot.corpus.cln)
# Create term document matrix
whynot.tdm <- TermDocumentMatrix(whynot.corpus.cln)

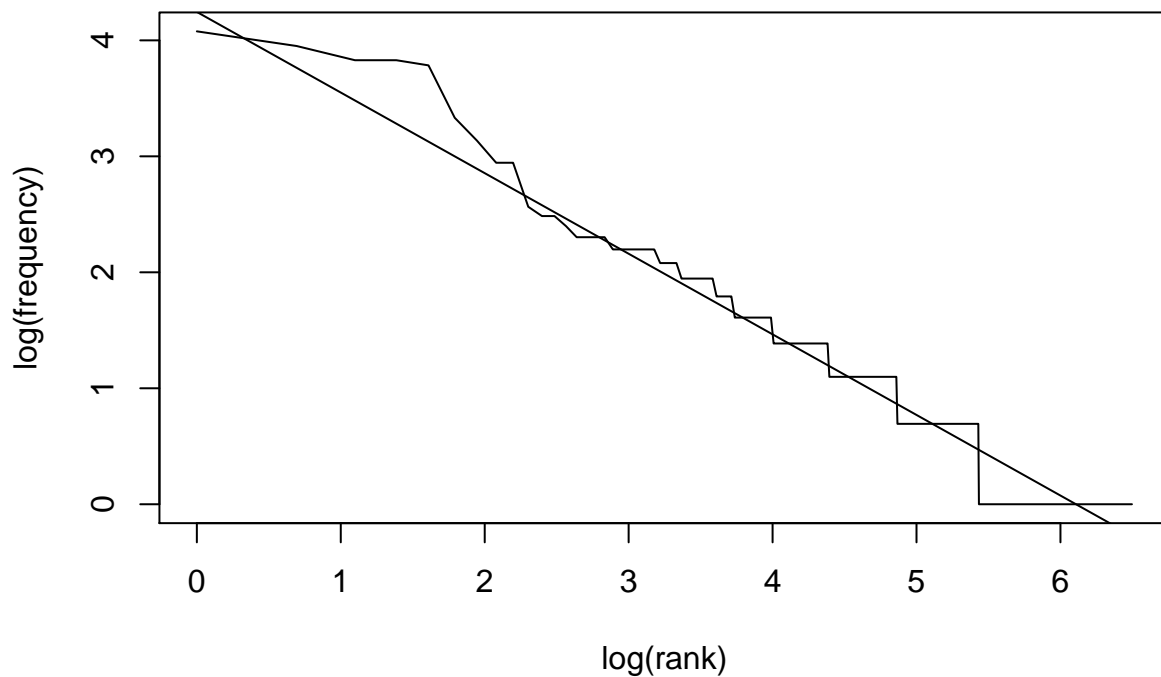
```

Create Zipf's plot

```

# Create Zipf's plot
Zipf_plot(whynot.tdm)

```

```
## (Intercept)          x
##  4.2448845 -0.6950045
```

Find terms that appear at least 10 times

```
# Find terms that occur at least 10 times
findFreqTerms(whynot.dtm, 10)
```

```
## [1] "treatment" "clinical" "needed" "recurrence" "trial"
## [6] "tumor" "diagnosis" "trials" "radiation" "standard"
## [11] "available" "time" "type" "doctor" "surgery"
## [16] "options" "qualify"
```

Data Visualization Create a wordcloud

```
# Create wordcloud with the most commonly used words larger and in the center
wordcloud(whynot.corpus.cln, random.order = FALSE)
```



Text Clustering

```
# Remove empty documents
# Code provided by agstudy
row.totals <- apply(whynot.dtm, 1, sum)
whynot.dtm <- whynot.dtm[row.totals > 0, ]
# Create LDA model with 7 topics
whynot.lda <- LDA(whynot.dtm, k = 7)
# Review 7 terms associated with each topic
term <- terms(whynot.lda, 7)
(term <- apply(term, MARGIN = 2, paste, collapse = ", "))
```

```
## Topic 1
## "tumor, qualify, one, point, doctors, time, yet"
## Topic 2
## "treatment, time, clinical, trials, trial, needed, standard"
## Topic 3
## "tumor, doctor, surgery, said, removed, recommended, time"
## Topic 4
## "trials, clinical, available, current, options, trial, informed"
## Topic 5
## "trial, offered, clinical, tumor, qualify, trials, treated"
## Topic 6
## "treatment, standard, trial, care, plan, clinical, working"
## Topic 7
## "clinical, trial, need, meningioma, surgery, patients, trials"
```

Sentiment analysis

```

# Perform sentiment analysis
whynot.sentiment <- analyzeSentiment(whynot.corpus.cln)
# Check the output of the sentiment analysis
head(whynot.sentiment)

```

```

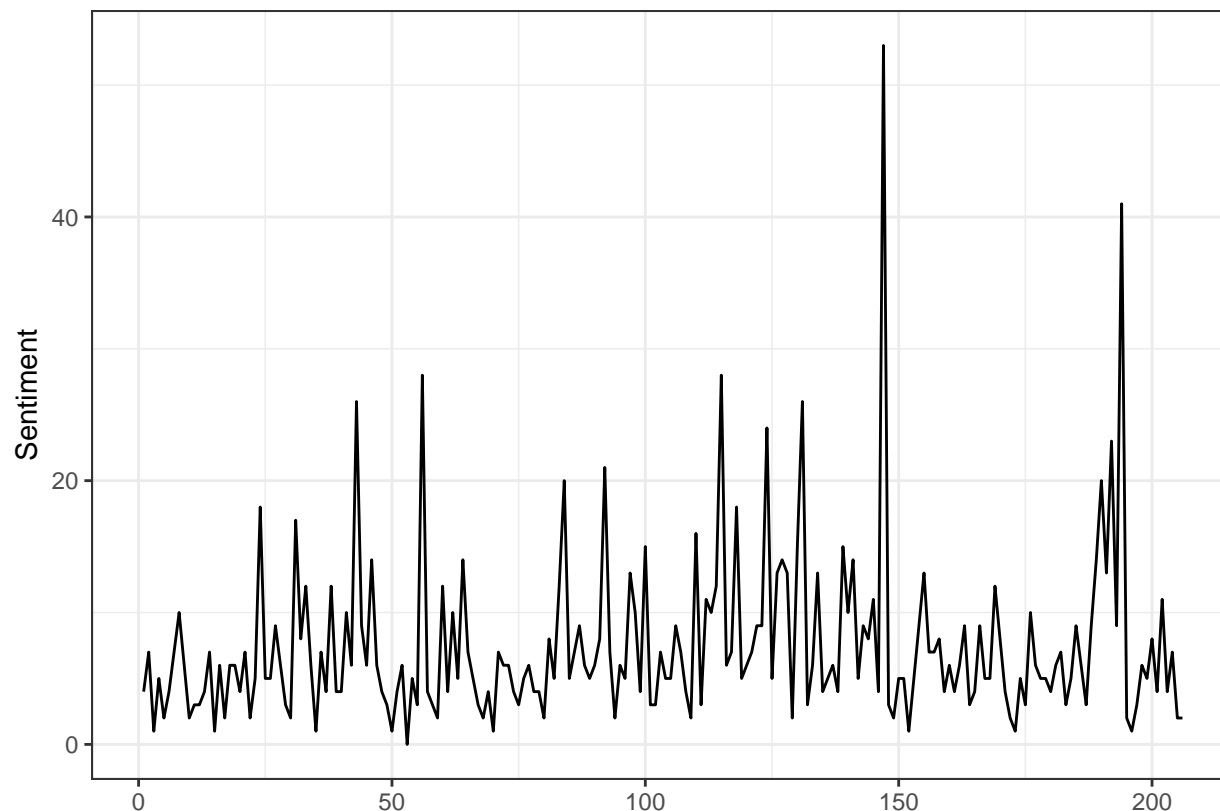
##      WordCount SentimentGI NegativityGI PositivityGI SentimentHE NegativityHE
## 1           4    0.2500000    0.0000000         0.25         0.0          0
## 2           7   -0.1428571    0.1428571         0.00         0.0          0
## 3           1    1.0000000    0.0000000         1.00         0.0          0
## 4           5    0.2000000    0.0000000         0.20         0.2          0
## 5           2   -0.5000000    0.5000000         0.00         0.0          0
## 6           4    0.2500000    0.0000000         0.25         0.0          0
##      PositivityHE SentimentLM NegativityLM PositivityLM RatioUncertaintyLM
## 1           0.0          0.0          0          0.0          0.25
## 2           0.0          0.0          0          0.0          0.00
## 3           0.0         -1.0          1          0.0          0.00
## 4           0.2          0.2          0          0.2          0.20
## 5           0.0          0.0          0          0.0          0.00
## 6           0.0          0.0          0          0.0          0.00
##      SentimentQDAP NegativityQDAP PositivityQDAP
## 1      0.2500000    0.0000000         0.25
## 2     -0.1428571    0.1428571         0.00
## 3      0.0000000    0.0000000         0.00
## 4      0.2000000    0.0000000         0.20
## 5      0.0000000    0.0000000         0.00
## 6      0.2500000    0.0000000         0.25

```

```

# Visualize the results of the sentiment analysis
plotSentiment(whynot.sentiment)

```



```
# Get summary results from sentiment analysis
summary(whynot.sentiment)
```

```
##      WordCount      SentimentGI      NegativityGI      PositivityGI
##  Min.   : 0.000    Min.   :-1.00000    Min.   :0.00000    Min.   :0.0000
## 1st Qu.: 4.000    1st Qu.: 0.00000    1st Qu.:0.00000    1st Qu.:0.0000
##  Median : 6.000    Median : 0.05556    Median :0.00000    Median :0.1667
##  Mean   : 7.461    Mean   : 0.11749    Mean   :0.09079    Mean   :0.2083
## 3rd Qu.: 9.000    3rd Qu.: 0.25000    3rd Qu.:0.16667    3rd Qu.:0.3077
##  Max.   :53.000    Max.   : 1.00000    Max.   :1.00000    Max.   :1.0000
##                      NA's   :1          NA's   :1          NA's   :1
##      SentimentHE      NegativityHE      PositivityHE
##  Min.   :-0.20000    Min.   :0.000000    Min.   :0.00000
## 1st Qu.: 0.00000    1st Qu.:0.000000    1st Qu.:0.00000
##  Median : 0.00000    Median :0.000000    Median :0.00000
##  Mean   : 0.01635    Mean   :0.005027    Mean   :0.02138
## 3rd Qu.: 0.00000    3rd Qu.:0.000000    3rd Qu.:0.00000
##  Max.   : 0.50000    Max.   :0.200000    Max.   :0.50000
##  NA's   :1          NA's   :1          NA's   :1
##      SentimentLM      NegativityLM      PositivityLM
##  Min.   :-1.000000    Min.   :0.00000    Min.   :0.00000
## 1st Qu.: 0.000000    1st Qu.:0.00000    1st Qu.:0.00000
##  Median : 0.000000    Median :0.00000    Median :0.00000
##  Mean   : 0.007715    Mean   :0.02974    Mean   :0.03745
## 3rd Qu.: 0.000000    3rd Qu.:0.00000    3rd Qu.:0.00000
##  Max.   : 1.000000    Max.   :1.00000    Max.   :1.00000
```

```
## NA's :1      NA's :1      NA's :1      NA's :1
## RatioUncertaintyLM SentimentQDAP      NegativityQDAP      PositivityQDAP
## Min. :0.00000      Min. : -0.50000      Min. :0.0000      Min. :0.0000
## 1st Qu.:0.00000      1st Qu.: 0.00000      1st Qu.:0.0000      1st Qu.:0.0000
## Median :0.00000      Median : 0.04167      Median :0.0000      Median :0.1111
## Mean :0.01871      Mean : 0.10690      Mean :0.0474      Mean :0.1543
## 3rd Qu.:0.00000      3rd Qu.: 0.22222      3rd Qu.:0.0000      3rd Qu.:0.2500
## Max. :0.66667      Max. : 1.00000      Max. :0.5000      Max. :1.0000
## NA's :1      NA's :1      NA's :1      NA's :1
```

Create n-grams

```
# Convert corpus to strings
whynot.string <- as.character(whynot.corpus.cln)[1]
# Separate words
whynot.words <- strsplit(whynot.string, " ", fixed = TRUE) [[1]]
# create bigrams
whynot.bi <- vapply(ngrams(whynot.words, 2), paste, "", collapse = " ")
# Check bigrams
whynot.bi[1:3]
```

```
## [1] "c(\" treatment\" \"treatment plan\" \"plan appeared\"
```

Text Clustering with n-grams

```
# Create bigram DTM
whynot.bi.dtm <- DocumentTermMatrix(whynot.corpus.cln,
                                     control = list(tokenize = BigramTokenizer))

# Remove empty terms
# Code provided by agstudy
bi.row.totals <- apply(whynot.bi.dtm, 1, sum)
whynot.bi.dtm <- whynot.bi.dtm[bi.row.totals > 0, ]
# Create bigram LDA model
whynot.bigram.lda <- LDA(whynot.bi.dtm, k = 5)
# Review 7 terms associated with each topic
not.term.bi <- terms(whynot.bigram.lda, 7)
(not.term.bi <- apply(not.term.bi, MARGIN = 2, paste, collapse = ", "))
```

```
##                                     Topic 1
##      "treatment, standard, clinical, options, tumor, trial, trials"
##                                     Topic 2
##      "tumor, doctor, surgery, said, removed, trial, recommended"
##                                     Topic 3
##      "trials, clinical, current, qualify, informed, diagnosed, condition"
##                                     Topic 4
##      "time, tumor, trial, clinical, need, diagnosis, needed"
##                                     Topic 5
##      "clinical, trial, trials, available, one, meningioma, find"
```

K-means clustering

```
# Create new DTM with TF-IDF
whynot.tfidf <- DocumentTermMatrix(whynot.corpus.cln, control = list(weighting = weightTfIdf))

## Warning in weighting(x): empty document(s): 53

# Check dimensions of DTM
dim(whynot.tfidf)
```

```
## [1] 206 663
# Run k-means with 3 clusters
whynot.kmeans <- kmeans(whynot.tfidf, 3, iter.max = 100)
table(whynot.kmeans$cluster)

##
## 1 2 3
## 2 3 201
# Identify ten words associated with each cluster
for (i in 1:3) {
  cat(paste("Cluster ", i, ": ", sep = ""))
  s <- sort(whynot.kmeans$centers[i, ], decreasing = TRUE)
  cat(names(s)[1:10], "\n")
}

## Cluster 1: know appeared plan sufficient treatment 100 clinical gross needed recurrence
## Cluster 2: want yet scare gotten appeared plan sufficient treatment 100 clinical
## Cluster 3: treatment time tumor trial clinical trials need surgery doctor standard
```

References agstudy. (2012). Remove empty documents from DocumentTermMatrix in R topicmodels? Retrieved from <https://stackoverflow.com/questions/13944252/remove-empty-documents-from-documenttermmatrix-in-r-topicmodels>

Bhalla, D. R: Keep/drop columns from data frame. Retrieved from <http://www.listendata.com/2015/06/r-keep-drop-columns-from-data-frame.html>

Feuerriegel, S. SentimentAnalysis. Retrieved from <https://www.rdocumentation.org/packages/SentimentAnalysis/versions/1.3-0>

Spell Checker for R. . . qdap::check_spelling. (2014). Retrieved from https://trinkerrstuff.wordpress.com/2014/09/04/spell-checker-for-r-qdapcheck_spelling/

techno@admin+. (2016). How to clean the twitter data using R-Twitter mining tutorial. Retrieved from <http://technokarak.com/how-to-clean-the-twitter-data-using-r-twitter-mining-tutorial.html>

Xie, Y. (2012). R - Markdown avoiding package loading messages. Retrieved from <https://stackoverflow.com/questions/13090838/r-markdown-avoiding-package-loading-messages>

Zhao, Y. (2016). Twitter data analysis with R - Text mining and social network analysis. Retrieved from <http://www.rdatamining.com/docs/twitter-analysis-with-r>